

**Quantitative Evaluierung spleißregulatorischer Motive  
in 5'-Spleißstellennähe im humanen Zellmodell**

**D I S S E R T A T I O N**

zur Erlangung des akademischen Grades  
Doctor rerum naturalium (*Dr. rer. nat.*)  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von  
Dipl.-Ing. (FH) Frank Schumann

Kiel, im Oktober 2013

Erstgutachter: Prof. Dr. Karin Krupinska

Zweitgutachter: Prof. Dr. Jochen Hampe

Tag der mündlichen Prüfung: 12.12.2013

Druckgenehmigung: 12.12.2013

*Meinen Eltern und  
meiner Freundin*

---

## Inhaltsverzeichnis

Abbildungsverzeichnis .....	IV
Tabellenverzeichnis .....	V
Abkürzungsverzeichnis .....	VI
<b>1 Zusammenfassung.....</b>	<b>1</b>
<b>2 Abstract.....</b>	<b>2</b>
<b>3 Einleitung.....</b>	<b>3</b>
3.1 Spleißen eukaryotischer Primärtranskripte im Kontext der Genexpression .....	3
3.2 Konstitutives Spleißen.....	4
3.2.1 Spleißreaktion und kanonische Sequenzelemente .....	4
3.2.2 Assemblierung des Spleißosoms .....	5
3.3 Alternatives Spleißen.....	6
3.4 Mechanismus der Spleißregulation .....	7
3.4.1 Intron- und Exon-Erkennungsmodelle .....	8
3.4.2 Spleißregulatoren.....	8
3.5 Bedeutung der 5'-Spleißstelle im Krankheitskontext .....	10
3.6 Programme zur Vorhersage von <i>cis</i> -Elementen und <i>trans</i> -Faktoren .....	11
3.7 Zielstellung .....	12
<b>4 Material und Methoden.....</b>	<b>14</b>
4.1 Chemikalien.....	14
4.2 Herstellung rekombinanter Plasmide .....	14
4.2.1 Polymerase-Ketten-Reaktion (PCR).....	14
4.2.2 Agarose-Gelelektrophorese.....	16
4.2.3 Aufreinigung von DNA-Fragmenten.....	17
4.2.4 Konzentrationsbestimmung von Nukleinsäuren.....	17
4.2.5 Kommerzielle Vektorsysteme .....	17
4.2.6 DNA-Restriktion .....	18
4.2.7 Ligation und Zirkularisierung von linearer DNA.....	18
4.2.8 Transformation von Plasmid-DNA in kompetente <i>E.coli</i> -Zellen.....	19
4.2.9 Isolierung von Plasmid-DNA aus <i>E.coli</i> .....	19
4.2.10 Sanger-Sequenzierung .....	20
4.3 Zellkultur.....	20
4.3.1 Kultivierung von eukaryotischen Zellen .....	20
4.3.2 Zellernte und Zellzahlbestimmung.....	21
4.3.3 Transfektion von eukaryotischen Zellen .....	22

---

<b>4.4</b>	<b>DNA- und RNA-Analysen .....</b>	<b>22</b>
4.4.1	Isolierung von Gesamt-DNA aus eukaryotischen Zellen .....	22
4.4.2	Quantitative Real-Time-PCR (qRT-PCR) .....	22
4.4.3	Isolierung von Gesamt-RNA aus eukaryotischen Zellen .....	23
4.4.4	cDNA-Synthese und Reverse-Transkriptase-PCR (RT-PCR) .....	24
4.4.5	Kapillarelektrophorese-laserinduzierter Fluoreszenz (CE-LIF) .....	24
4.4.6	Illumina-Tiefensequenzierung .....	25
<b>4.5</b>	<b>Proteinanalysen .....</b>	<b>26</b>
4.5.1	Fluoreszenzaktivierte Zellsortierung (FACS) .....	26
4.5.2	Fluoreszenzmikroskopie .....	27
<b>4.6</b>	<b>Biometrische Datenauswertung .....</b>	<b>28</b>
<b>5</b>	<b>Ergebnisse .....</b>	<b>29</b>
<b>5.1</b>	<b>Das Spleißreportersystem .....</b>	<b>29</b>
5.1.1	Klonierung und Funktionalität des Reporters .....	29
5.1.2	Funktionelle Reportervalidierung .....	30
5.1.3	Identifizierte spleißregulatorische Loci .....	32
5.1.4	Bestimmung der Transfektionseffizienz .....	38
<b>5.2</b>	<b>Etablierung einer randomisierten Mutagenese .....</b>	<b>39</b>
5.2.1	Herstellung randomisierter Mutagenesebanken .....	39
5.2.2	Anreicherung der randomisierten Mutagenesebank .....	40
<b>5.3</b>	<b>Identifizierung von spleißregulatorischen Motiven (SRMs) .....</b>	<b>41</b>
5.3.1	Sequenzabgleich der Mutagenesebanken .....	42
5.3.2	Motivabdeckung der transfizierten Mutagenesebanken .....	42
5.3.3	Sequenzabgleich der Transkript-Pools .....	43
5.3.4	Häufigkeitsverteilungen der Motive .....	44
5.3.5	Aufklärung des Mischeffektes in den Transkriptdaten .....	46
5.3.6	Berechnung des skalierten Anreicherungsindex ( $AI_{sk}$ ) .....	47
5.3.7	Identifizierung von Enhancer- und Silencer-Sequenzen .....	48
<b>5.4</b>	<b>Validierung der identifizierten SRMs .....</b>	<b>50</b>
<b>5.5</b>	<b>Charakterisierung der identifizierten SRMs .....</b>	<b>50</b>
5.5.1	Zelllinienübergreifender Vergleich der SRMs .....	50
5.5.2	Sequenzhomologien von SRMs .....	52
5.5.3	Analyse der Bindestelle der Spleißosomkomponente U1-snRNP .....	53
5.5.4	Annotation der SRMs .....	53
<b>6</b>	<b>Diskussion .....</b>	<b>56</b>
<b>6.1</b>	<b>Etablierte Methode zur quantitativen Evaluierung von SRMs .....</b>	<b>56</b>
<b>6.2</b>	<b>Relevanz der identifizierten SRMs .....</b>	<b>58</b>
<b>6.3</b>	<b>Relevanz weiterer spleißregulatorischer Faktoren .....</b>	<b>60</b>
<b>6.4</b>	<b>Gezielter Abbau der Primärtranskripte .....</b>	<b>61</b>
<b>6.5</b>	<b>Spleißrelevante Mutationen im humanen Genom .....</b>	<b>62</b>
<b>6.6</b>	<b>Ausblick .....</b>	<b>63</b>

---

<b>Literaturverzeichnis .....</b>	<b>65</b>
<b>Anhang.....</b>	<b>74</b>
<b>Erklärung .....</b>	<b>116</b>
<b>Lebenslauf .....</b>	<b>117</b>
<b>Veröffentlichungen.....</b>	<b>118</b>
<b>Danksagung.....</b>	<b>119</b>

## Abbildungsverzeichnis

Abb. 1: Schematische Darstellung des prä-mRNA-Spleißens.....	4
Abb. 2: Schematische Darstellung der Spleißosom-Assemblierung.....	6
Abb. 3: Basismodelle des alternativen Spleißens .....	7
Abb. 4: Intron- und Exon-Erkennungsmodelle .....	8
Abb. 5: Schematische Darstellung einer Spleißregulation .....	10
Abb. 6: Aufbau der Expressionsvektoren pAcGFP1-N1 und pDsRed-Express-N1.....	18
Abb. 7: Schematische Darstellung des Forward-Primers zur Amplifikation des Sequenzausschnitts der Tiefensequenzierung .....	26
Abb. 8: Klonierungsschema des rHBBg-Reporters .....	29
Abb. 9: Schematische Darstellung der Funktionalität des Spleißreportersystems.....	30
Abb. 10: Funktionelle Validierung des Spleißreportersystems auf Transkriptebene.....	31
Abb. 11: Funktionelle Validierung des Spleißreportersystems auf Proteinebene in HeLa- Zellen.....	32
Abb. 12: Bioinformatisch identifizierte spleißregulatorische Loci.....	33
Abb. 13: CE-LIF-basierte Bewertung der bioinformatisch identifizierten spleißregulatorischen <i>HBB</i> -Loci .....	35
Abb. 14: FACS-basierte Bewertung der bioinformatisch identifizierten spleißregulatorischen <i>HBB</i> -Loci .....	36
Abb. 15: Motivkomplexität .....	37
Abb. 16: Plasmidstandardkurve zur Berechnung der Plasmidkopienzahl .....	38
Abb. 17: Relative Zusammensetzung der DL- und PL-Mutagenesebanken .....	42
Abb. 18: Relative Zusammensetzung des DL-Transkript-Pools .....	43
Abb. 19: Häufigkeitsverteilung der Motive in der DL-Mutagenesebank .....	44
Abb. 20: Häufigkeitsverteilung der Motive in den DL-Transkript-Pools.....	45
Abb. 21: Identifizierung eines NMD-Effektes im DL- und PL-Transkript-Pool in HeLa-Zellen (Replik 1) .....	46
Abb. 22: Quantifizierung des NMD-Effektes in Abhängigkeit von Locus und Zelltyp .....	47
Abb. 23: Schematische Darstellung der Berechnung des $AI_{sk}$ des DL-Reporterbereichs in HeLa-Zellen (Replik 1) .....	48
Abb. 24: Identifizierte Enhancer- und Silencer-Sequenzen des DL-Reporterbereichs in HeLa- Zellen.....	49
Abb. 25: Validierung ausgewählter SRMs des DL-Reporterbereichs in HeLa-Zellen .....	50
Abb. 26: Schnittmengendiagramme des zelllinienübergreifenden Vergleichs der SRMs.....	51
Abb. 27: Clusteranalyse der SRMs des DL-Reporterbereichs .....	52
Abb. 28: Analyse der Bindestelle des U1-snRNPs.....	53
Abb. 29: Annotation der Enhancer- und Silencer-Sequenzen .....	54
Abb. 30: Schematische Darstellung der Interaktionen von spleißregulatorischen Motiven und Bindepateern zur Erkennung einer 5'-Spleißstelle .....	59

---

## Tabellenverzeichnis

Tab. 1: Vorhersageprogramme von <i>cis</i> -Elementen und <i>trans</i> -Faktoren .....	12
Tab. 2: Berechnete Plasmid- und Genkopienzahlen pro diploider HeLa-Zelle .....	38
Tab. 3: Mutageneseeffizienzen der PCR-Ansätze zur Herstellung einer Mutagenesebank.....	40
Tab. 4: Vergleich der Methoden zur Anreicherung von Mutagenesebanken .....	41
Tab. 5: Anzahlen identifizierter Enhancer- und Silencer-Sequenzen in Abhängigkeit von Locus und Zelltyp.....	49
Tab. 6: Gemeinsame DL- und PL-Motive der Zelllinien HeLa, HuH7 und SH-SY5Y .....	51



## Abkürzungsverzeichnis

### • Allgemeine Abkürzungen

Abb.	Abbildung
Abschn.	Abschnitt
AI	Anreicherungsindex
AI <sub>sk</sub>	skalierter Anreicherungsindex
ATP	Adenosintriphosphat
bp	Basenpaar(e)
cDNA	engl. <i>complementary DNA</i> (komplementäre DNA)
CE-LIF	engl. <i>capillary electrophoresis with laser-induced fluorescence</i> (fluoreszenzbasierte Kapillarelektrophorese)
DL	distaler Locus
DNA	engl. <i>deoxyribonucleic acid</i> (Desoxyribonukleinsäure)
dNTP	Desoxy-Ribonukleosidtriphosphat
E <sub>ein</sub>	engl. <i>exon inclusion</i> (Exon-Einschluss)
E <sub>aus</sub>	engl. <i>exon skipping</i> (Exon-Ausschluss)
EJC	engl. <i>exon junction complex</i> (Exon-Junction-Komplex)
ESE	engl. <i>exonic splicing enhancer</i> (exonischer Spleiß-Enhancer)
ESS	engl. <i>exonic splicing silencer</i> (exonischer Spleiß-Silencer)
FACS	engl. <i>fluorescence activated cell sorting</i> (fluoreszenzaktivierte Zellsortierung)
g	Erdgravitation
GFP	grün fluoreszierendes Protein
hnRNP	engl. <i>heterogeneous nuclear ribonucleoprotein</i> (heterogenes nukleäres Ribonukleoprotein)
Indel	Insertion bzw. Deletion
IQR	engl. <i>interquartile range</i> (Interquartilsabstand)
ISE	engl. <i>intronic splicing enhancer</i> (intronischer Spleiß-Enhancer)
ISS	engl. <i>intronic splicing silencer</i> (intronischer Spleiß-Silencer)
KS-Test	Kolmogorow-Smirnow-Test
L <sub>xbisy</sub>	Locus der Positionen x bis y
MCS	engl. <i>multiple cloning site</i> (multiple Klonierungsstelle)
mRNA	engl. <i>messenger RNA</i> (Boten-RNA)
Mu	Mutante
MW	Molekulargewicht
NMD	engl. <i>nonsense-mediated mRNA decay</i> (Nonsens-vermittelter mRNA-Abbau)
nt	Nukleotid(e)
OD	optische Dichte
ORF	engl. <i>open reading frame</i> (offener Leserahmen)
PCR	engl. <i>polymerase chain reaction</i> (Polymerase-Ketten-Reaktion)
PL	proximaler Locus
pPTC	engl. <i>putative premature termination codon</i> (putatives vorzeitiges Translationsstoppkodon)
PTC	engl. <i>premature termination codon</i> (vorzeitiges Translationsstoppkodon)

---

qRT-PCR	engl. <i>quantitative real-time polymerase chain reaction</i> (quantitative Echtzeit- Polymerase-Ketten-Reaktion)
RFP	rot fluoreszierendes Protein
rHBBg	RFP-HBB-GFP
RNA	engl. <i>ribonucleic acid</i> (Ribonukleinsäure)
RRM	engl. <i>RNA recognition motif</i> (RNA-Erkennungsmotiv)
RS-Domäne	Arginin/Serin-reiche Domäne
RT-PCR	engl. <i>reverse transcriptase polymerase chain reaction</i> (reverse Transkriptase-Polymerase-Ketten-Reaktion)
RUST	engl. <i>regulated unproductive splicing and translation</i> (unproduktives Spleißen und regulierte Translation)
siRNA	engl. <i>small interfering RNA</i> (kurze interferierende RNA)
SNP	engl. <i>single nucleotide polymorphism</i> (Einzelnukleotid-Polymorphismus)
snRNP	engl. <i>small nuclear ribonucleoprotein particle</i> (kleines nukleäres Ribonukleoprotein-Partikel)
SRM	spleißregulatorisches Motiv
Tab.	Tabelle
U-Test	Wilcoxon-Mann-Whitney-Test
UTR	engl. <i>untranslated region</i> (untranslatierte Region)
U-snRNA	engl. <i>uridine-rich small nuclear RNA</i> (Uridin-reiche kleine nukleäre RNA)
Vol. %	Volumenprozent
v/v	Volumenkonzentration einer Substanz in einer Lösung
WA	Abweichung zum Wildtyp
WT	Wildtyp
w/v	Massenkonzentration einer Substanz in einer Lösung

#### • IUPAC-Nukleotidcode

A	Adenin
C	Cytosin
G	Guanin
T	Thymin
U	Uracil
R	A oder G
Y	C oder T
S	G oder C
W	A oder T
K	G oder T
M	A oder C
B	C, G oder T
D	A, G oder T
H	A, C oder T
V	A, C oder G
N	jede Base

## 1 ZUSAMMENFASSUNG

Im Rahmen der vorliegenden Arbeit wurden experimentell spleißregulatorische Motive in der Nähe einer 5'-Spleißstelle in drei verschiedenen Zellkulturmodellen quantitativ evaluiert und entsprechend ihres Einflusses auf die Spleißaktivität klassifiziert. Insgesamt konnten 1812 Enhancer- und 2803 Silencer-Sequenzen am Beispiel des humanen  $\beta$ -Globingens (*HBB*) identifiziert werden.

Über 60 % der krankheitsassoziierten Mutationen im humanen Genom verursachen fehlerhafte Spleißreaktionen. Mit einer 50 %igen Wahrscheinlichkeit sind die Mutationen in exonischen spleißregulatorischen Loci lokalisiert. Diese Loci befinden sich hauptsächlich bis zu 20 bp stromaufwärts (5') und stromabwärts (3') der Exon-Intron-Grenzen und haben in der Nähe einer 5'-Spleißstelle den stärksten Einfluss auf die Spleißaktivität. Um zukünftig die Spleißrelevanz von Mutationen im humanen Genom vorhersagen zu können, war es das Ziel dieser Arbeit, neue spleißregulatorische Motive in 5'-Spleißstellennähe zu identifizieren.

Zunächst wurde dafür ein Plasmid konstruiert, das den vollständigen Genkontext des Kandidatengens *HBB* beinhaltet und fluoreszenzbasiert den Vergleich konstitutiver und alternativer Spleißereignisse in eukaryotischen Zellen ermöglichte. In Zellkulturexperimenten konnten die Loci -19 bis -12 (distal) und -7 bis -1 (proximal) der 5'-Spleißstelle des *HBB*-Introns 2 als regulatorische Bereiche identifiziert werden. Die randomisierte Mutagenese der beiden Loci und die transiente Transfektion der resultierenden Mutagenesebanken in die epitheliale Zelllinie HeLa, in die Leberzelllinie HuH7 und in die neuronale Zelllinie SH-SY5Y, erfolgten im Anschluss. Durch die Quantifizierung der konstitutiv gespleißten Transkripte mittels Tiefensequenzierung wurde die relative Spleißaktivität jedes Motivs bestimmt.

Unabhängig von der Zelllinie konnten 1812 Enhancer-Sequenzen identifiziert werden, davon 1301 Sequenzen distal und 511 Sequenzen proximal der 5'-Spleißstelle. Von den 2803 Silencer-Sequenzen wurden 2037 Sequenzen distal und 766 Sequenzen proximal der 5'-Spleißstelle identifiziert. Ein zelllinienübergreifender Motivvergleich ergab, dass in HeLa-Zellen die meisten Motive konstitutiv transkribiert werden, während die Zelllinie SH-SY5Y die stärkste alternative Spleißaktivität aufweist. In abschließenden Assoziationsstudien zur Identifizierung der Bindepartner der neuen, experimentell ermittelten spleißregulatorischen Motive, zeigte sich, dass Spleißaktivatoren vorwiegend im proximalen Locus und Spleißrepressoren überwiegend im distalen Locus binden. Dabei fungieren vor allem die Spleißfaktoren hnRNP A/B und YB1 als Aktivatoren und hnRNP H/F und CUG-BP als Repressoren.

Die Ergebnisse dieser Arbeit dienen als Ausgangspunkt zur zelltypabhängigen Aufklärung krankheitsbezogener Mutationen in spleißregulatorischen Regionen im humanen Genom.

## 2 ABSTRACT

In the present work splicing regulatory motifs near the 5' splice site were quantitatively evaluated and classified according to their impact on splicing activation in three different cell culture models. Altogether 1812 enhancer sequences and 2803 silencer sequences were identified using the example of the human  $\beta$ -globin gene (*HBB*).

More than 60 % of disease-associated mutations in the human genome mediate an inaccurate splicing signal. With a feasibility of 50 % these mutations are located in exonic splicing regulatory loci. Especially the 5' splice site with 20 bp upstream (5') and downstream (3') has important regulatory function. To improve the prediction of splicing relevant mutations in the human genome, the aim of this work was the identification of new splicing regulatory motifs near the 5' splice site.

Based on transient transfection of a self-designed plasmid containing the whole genomic region of the *HBB* gene flanked by genes encoding for fluorescent proteins, constitutively and alternatively spliced *HBB* transcripts could be compared in eukaryotic cells. In cell culture experiments the loci -19 to -12 (distal) and -7 to -1 (proximal) of the 5' splice site of the *HBB* intron 2 were identified as regulatory regions. Afterwards, these regions were randomly mutated and the resulting mutagenic libraries were transfected in the epithelial cell line HeLa, in the liver cell line HuH7, and in the neuronal cell line SH-SY5Y. The quantitative evaluation of constitutively spliced transcripts was done with second-generation sequencing. The relative splice activity of a given motif was calculated and the motifs were classified as splicing enhancer or silencer.

In total, 1812 enhancer sequences, thereof 1301 sequences distal and 511 sequences proximal to the 5' splice site, and 2803 silencer sequences, separated into 2037 distal and 766 proximal sequences to the 5' splice site, were identified. The comparison of the individual motifs revealed that in HeLa cells most motifs were spliced constitutively, while the SH-SY5Y cells showed the strongest alternative splicing activity. To further identify binding partners of the newly determined splicing regulatory motifs, association studies were done. These analysis showed that splicing activators preferentially bind proximal to the 5' splice site, whereas repressors predominantly bind to the distal region. In detail, the splicing factors hnRNP A/B and YB1 were identified as activators and hnRNP H/F as well as CUG-BP as repressor proteins.

In summary, the results of this work can be used to better identify disease-associated mutations in splice regulatory regions of the human genome in a cell type-dependent manner in the future.

### 3 EINLEITUNG

Die Sequenzierung des humanen Genoms zeigte, dass der Mensch, trotz einer Genomgröße von 3 Milliarden Basenpaaren, nur über 20.000-25.000 proteinkodierende Gene verfügt (Venter et al. 2001; Lander et al. 2001; Hattori 2005; McKernan et al. 2009). Dies sind ähnlich viele Gene wie die Fruchtfliege *Drosophila melanogaster* besitzt und weniger als das Genom der Ackerschmalwand *Arabidopsis thaliana* aufweist (Adams 2000; Walbot 2000). Im Gegensatz zu diesen niederen Organismen steht der geringen Anzahl an menschlichen Genen ein bis zu fünfmal größeres Proteom von mehr als 100.000 Proteinen gegenüber (Anderson et al. 2001; Wortman et al. 2003; Lin et al. 2007). Innerhalb des letzten Jahrzehnts wurden eine Reihe von Mechanismen beschrieben, die die Entstehung funktionell verschiedener Proteine aus einem Gen erklären und das bis dahin geltende „Ein-Gen-ein-Protein“-Dogma revidieren. Hauptverantwortlich für die Diversifikation des Proteoms ist das alternative Spleißen (Maniatis & Tasic 2002). Neuesten Erkenntnissen zufolge werden ca. 95 % der humanen Multi-Exon-Gene alternativ gespleißt (Pan et al. 2008).

Die biologische Relevanz des RNA-Spleißens spiegelt sich jedoch nicht nur in der Proteom-Diversifikation, sondern auch in der Entwicklung bzw. dem Verlauf komplexer Erkrankungen wie Progerie (Hutchinson-Gilford-Syndrom), Sarkoidose und Spinale Muskelatrophie wider (Cartegni & Krainer 2002; Valentonyte et al. 2005; Scaffidi & Misteli 2006). Über die Hälfte der menschlichen Krankheiten, die durch Mutationen verursacht werden, können auf Fehlspleißen zurückgeführt werden (López-Bigas et al. 2005).

#### 3.1 Spleißen eukaryotischer Primärtranskripte im Kontext der Genexpression

Die Genexpression in Eukaryoten beginnt mit der Umschreibung einer spezifischen DNA-Sequenz in ein Primärtranskript (prä-mRNA) im Zellkern. Da die meisten eukaryotischen Gene mehrere proteinkodierende DNA-Sequenzen (Exons) besitzen, die von durchschnittlich zehn bis 20-mal längeren nicht-proteinkodierenden DNA-Abschnitten (Introns) separiert sind, wird die reife mRNA durch das Entfernen der Introns aus der prä-mRNA und dem Religieren der Exons prozessiert (Ast 2004). Dieser ko- und posttranskriptionell verlaufende nukleäre Prozess wird als Spleißen bezeichnet. Die Translation der mRNA in die spezifische Aminosäuresequenz erfolgt im Anschluss im Zytoplasma (Cooke & Alwine 2002).

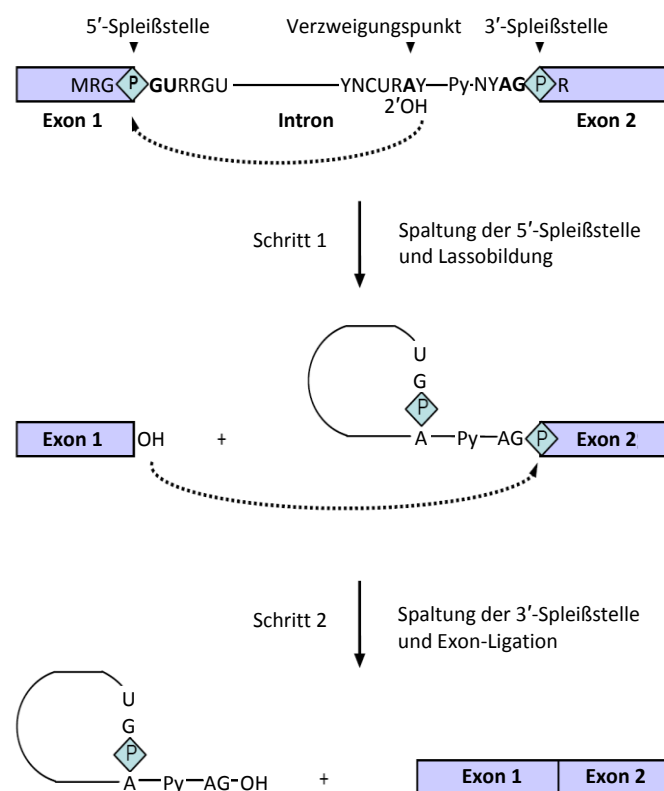
Zu den prä-mRNA-Modifikationen zählen neben dem Spleißen das Anfügen eines 7-Methyl-Guanosins am 5'-Ende (engl. *capping*) und eine Polyadenylierung am 3'-Ende. Die 5'-Cap-Struktur schützt die mRNA vor einem 5'-3'-exonukleolytischen Abbau und ist ein Interaktionspartner verschiedener Translationsinitiationsfaktoren. Die 3'-Polyadenylierung erhöht die mRNA-Stabilität und unterstützt ebenfalls die Initiation der Translation. Die prä-mRNA-

Modifikationen sind räumlich und zeitlich gekoppelt und hauptverantwortlich für die Regulation der Genexpression auf Transkriptebene (Shatkin & Manley 2000; Maniatis & Reed 2002; Proudfoot et al. 2002).

## 3.2 Konstitutives Spleißen

### 3.2.1 Spleißreaktion und kanonische Sequenzelemente

Die Introns bilden mit ca. 95 % den Hauptanteil der humanen prä-mRNA (Ast 2004). Das präzise Entfernen des hohen Intron-Anteils und das anschließende Religieren der Exons erfolgt in zwei nacheinander ablaufenden Transesterifizierungsreaktionen, welche von den vier hoch konservierten kanonischen prä-mRNA-Sequenzelementen, der 5'-Spleißstelle, der 3'-Spleißstelle, des Verzweigungspunktes und eines Polypyrimidintrakts variabler Länge vermittelt werden (Abb. 1).



**Abb. 1: Schematische Darstellung des prä-mRNA-Spleißens** (modifiziert nach Gouloudis 2004): Konservierte Nukleotide und Phosphate sind mit Pfeilen markiert. „Py“ symbolisiert einen Polypyrimidintrakt und die gestrichelten Linien repräsentieren die nukleophilen Angriffe der Hydroxylgruppen an den Exon-Intron-Grenzen.

Im ersten Umesterungsschritt wird über einen nukleophilen Angriff der 2'-OH-Gruppe des Adenosins des Verzweigungspunktes auf die Phosphodiesterbindung des Guanins der 5'-Spleißstelle (Spleißdonor) das 5'-Exon abgespalten. Der Verzweigungspunkt ist 18-40 nt stromaufwärts der 3'-Spleißstelle (Spleißakzeptor) lokalisiert (Krämer 1996; Cartegni et al. 2002).

Über eine 2',5'-Phosphodiesterbindung zwischen dem freien 5'-Ende des Introns und dem Adenosin des Verzweigungspunktes kommt es zur Ausbildung einer Lasso-ähnlichen Struktur (Lariat). Im zweiten Schritt greift die 3'-OH-Gruppe des freien 5'-Exons die Phosphodiesterbindung an der 3'-Spleißstelle an, wodurch die Freisetzung der Intron-Lassostruktur und die Ligation benachbarter Exons erfolgt (Krämer 1996; Staley & Guthrie 1998).

Erwähnenswert ist, dass knapp 1 % aller Introns GC statt GU als 5'-Spleißstelle besitzen (Burset et al. 2000; Thanaraj & Stamm 2003).

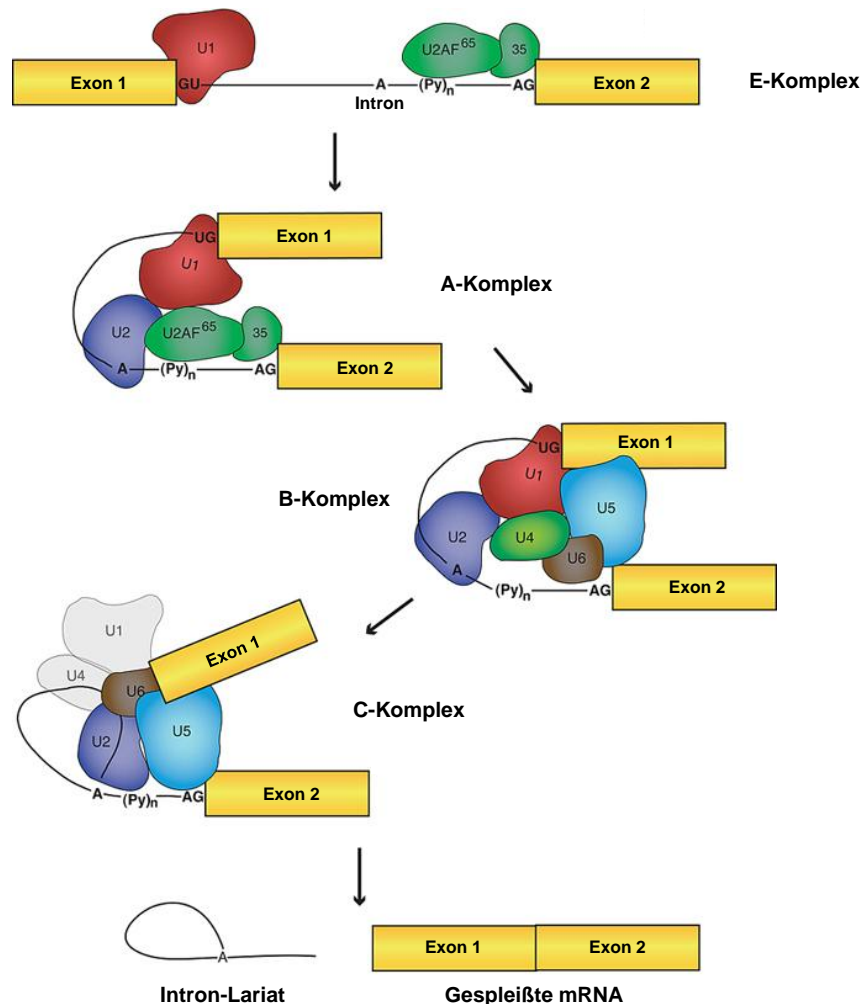
Neben den beschriebenen konstitutiven GU-AG-Introns, die mittels Hauptspleißosom prozessiert werden (U2-abhängiges Spleißen), existiert noch eine seltenere Form, die sogenannten AU-AC-Introns. Diese sind durch andere Konsensussequenzen gekennzeichnet und werden von einem Nebenspleißosom entfernt (U12-abhängiges Spleißen) (Levine & Durbin 2001).

### 3.2.2 Assemblierung des Spleißosoms

Die Spleißreaktion wird durch das Spleißosom, einem hoch dynamischen Ribonukleoprotein-Komplex, der seine strukturelle Zusammensetzung während der biochemischen Reaktion verändert, katalysiert (Staley & Guthrie 1998). An der Assemblierung des reifen Spleißosoms sind fünf kleine nukleäre Ribonukleoprotein-Partikel (engl. *small nuclear ribonucleoprotein particle*, snRNPs) (U1-, U2-, U4-, U5- und U6-snRNP) und zusätzliche Gruppen Spleißosom-assoziiierter Spleißfaktoren beteiligt. Jedes snRNP setzt sich aus einer Uridin-reichen kleinen nukleären RNA (engl. *uridine-rich small nuclear RNA*, U-snRNA) zusammen, an die sieben gemeinsame Proteine (Sm- und Sm-ähnliche (LSm) Proteine) fest gebunden sind und die mit zahlreichen weniger stabil assoziierten Spleißfaktoren interagieren (Bach et al. 1989).

Der erste Schritt der Spleißosom-Assemblierung ist die Ausbildung des E-Komplexes (engl. *early complex*) durch die Bindung des U1-snRNPs an die 5'-Spleißstelle der prä-mRNA. Der Spleißfaktor U2AF bindet den Polypyrimidintrakt sowie die 3'-Spleißstelle. Die darauffolgende Interaktion zwischen Spleißfaktor 1 (SF1) und U2AF führt zur Erkennung des Verzweigungspunktes (Hoffman & Grabowski 1992; Cunningham et al. 1995). Die Bildung des prä-Spleißkomplexes (A-Komplex) folgt auf eine ATP-abhängige Rekrutierung des U2-snRNPs an den Verzweigungspunkt. Aufgrund der daraus resultierenden Konformationsänderung wird das konservierte Adenosin der prä-mRNA in eine sterisch leicht zugängliche Position gebracht. Durch Integration des bereits assemblierten U4/U6-U5-tri-snRNP-Komplexes wird das präkatalytische Spleißosom (B-Komplex) gebildet. RNA-RNA-Wechselwirkungen veranlassen die Aufhebung der Basenpaarungen der U1-snRNA und der 5'-Spleißstelle sowie die Auflösung der intermolekularen Basenpaarungen zwischen den U4- und U6-snRNAs. Nach Verlust der U6-snRNA-Inhibition, die durch die U4-snRNA vermittelt wurde, entsteht das U2-U6-snRNA-katalytische Zentrum des Spleißosoms. Im Anschluss an die erste Umesterungsreaktion bildet

sich der C-Komplex. Im weiteren Verlauf werden die Exons 1 und 2 zu einer reifen mRNA verknüpft, das Intron-Lariat enzymatisch abgebaut und die dissoziierten Spleißkomponenten in weiteren Spleißzyklen verwendet (Chiara et al. 1997; Collins & Guthrie 2001; Will & Lührmann 2001). In Abbildung 2 ist die Assemblierung des Spleißosoms schematisch dargestellt.

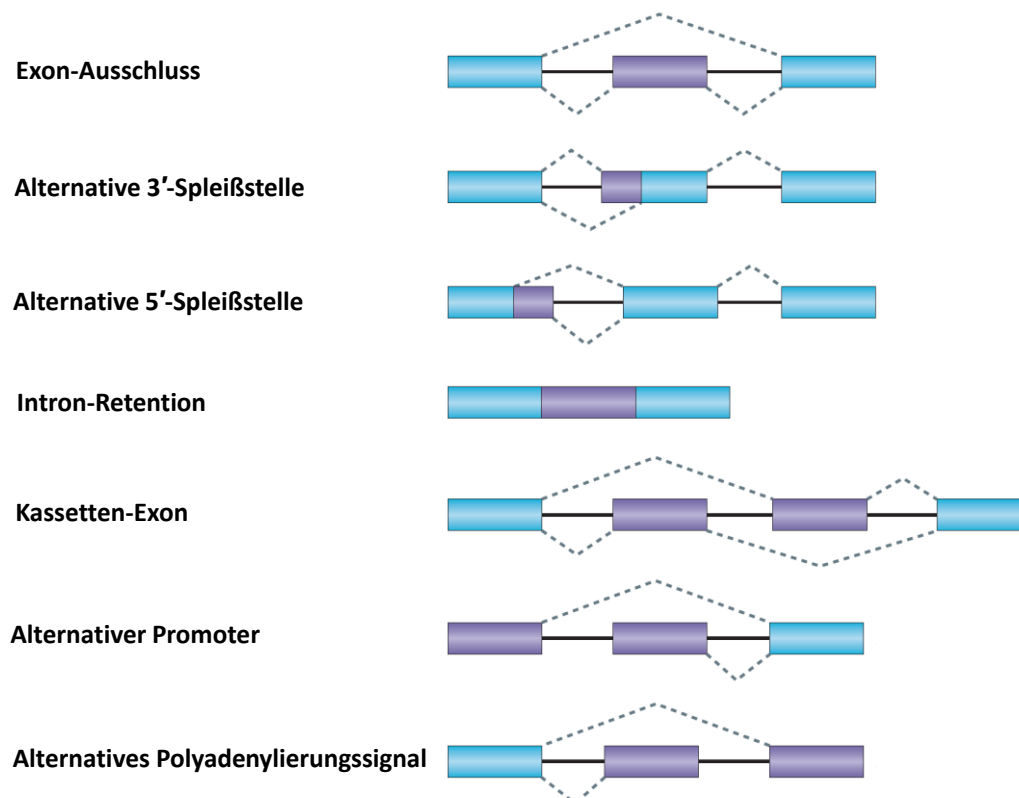


**Abb. 2: Schematische Darstellung der Spleißosom-Assemblierung** (modifiziert nach Solis et al. 2008): „GU“, „AG“, „A“ und „(Py)<sub>n</sub>“ repräsentieren die kanonischen Sequenzelemente 5'- und 3'-Spleißstelle, Verzweigungspunkt sowie Polypyrimidintrakt variabler Länge. Das Spleißosom besteht aus den fünf snRNPs „U1“, „U2“, „U4“, „U5“ und „U6“. Der Spleißfaktor U2AF setzt sich aus den Untereinheiten „U2AF<sup>65</sup>“ und „35“ zusammen.

### 3.3 Alternatives Spleißen

Neben der Verwendung konstitutiver Spleißstellen können auch alternative Spleißstellen in einer prä-mRNA genutzt werden. Durch die Wahl alternativer Spleißstellen durch das Spleißosom entstehen verschiedene mRNA-Isoformen, die häufig in funktionell verschiedene Proteine translatiert werden (Modrek et al. 2001). Alternatives Spleißen erfolgt in Abhängigkeit von Gewebe, Zelltyp, Entwicklungsstadium und infolge externer Stimuli, z.B. Stress, Hitzeschock oder Ionenkonzentrationen (Lopez 1998). Derzeit sind sieben Basismodelle des alternativen Spleißens bekannt (Abb. 3).





**Abb. 3: Basismodelle des alternativen Spleißens** (modifiziert nach Keren et al. 2010): Die blauen und violetten Rechtecke stellen konstitutive und alternative Exons dar, die durch Introns (schwarze Linien) getrennt sind. Die blauen gestrichelten Linien geben mögliche Spleißvorgänge an.

Eines der bemerkenswertesten Beispiele für alternatives Spleißen ist die Prozessierung des Gens *DSCAM* der Fruchtfliege *Drosophila melanogaster*. *DSCAM* kodiert ein Zelloberflächenprotein mit neuronaler Funktion und generiert aufgrund zahlreicher Kombinationen alternativer Spleißmodelle 38.016 Proteinisoformen. Dies sind mehr als doppelt so viele Isoformen wie vorhandene Gene im *Drosophila*-Genom (Neves et al. 2004).

In höheren Eukaryoten bildet der Exon-Ausschluss ( $E_{aus}$ ) mit 40 % den Hauptanteil an alternativen Spleißereignissen, während in weniger komplexen Organismen dieses Spleißmodell unterrepräsentiert ist (Sugnet et al. 2004; Alekseyenko et al. 2007).

Alternatives Spleißen kann in jeder Region der prä-mRNA erfolgen. Insertionen bzw. Deletionen im 5'-UTR-Bereich haben dabei Auswirkungen auf das Proteinexpressionslevel. Zu dem gleichen Ergebnis führen, aufgrund einer veränderten mRNA-Stabilität, alternative Spleißereignisse im 3'-UTR-Bereich. Im Gegensatz dazu hat eine alternative Spleißreaktion innerhalb der proteinkodierenden Regionen eine Veränderung der Proteinstrukturen und -funktionen zur Folge (Cartegni et al. 2002; Keren et al. 2010).

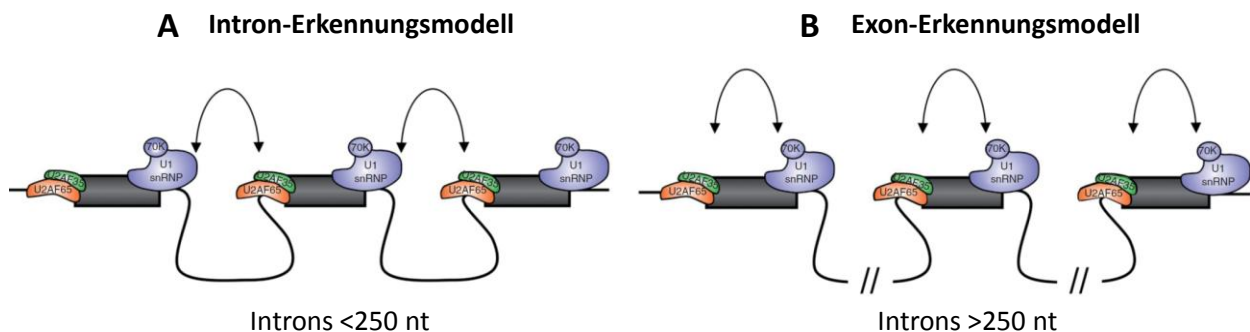
### 3.4 Mechanismus der Spleißregulation

Die Erkennung einer Spleißstelle ist ein komplexer Mechanismus, der von vielen intrinsischen und extrinsischen Faktoren bestimmt wird. Das Zusammenspiel dieser Spleißfaktoren wird als

Spleißcode bezeichnet, welcher aktuell nur lückenhaft entschlüsselt ist (Wang & Cooper 2007; Barash et al. 2010).

### 3.4.1 Intron- und Exon-Erkennungsmodelle

Die Exon/Intron-Architektur der eukaryotischen prä-mRNA bildet die Basis der Spleißstellen-erkennung, wobei hier hauptsächlich die Länge der Introns und Exons von Bedeutung ist. In Abhängigkeit der Intron- und Exon-Längen findet die Spleißstellenerkennung durch die Kommunikation zwischen dem U1-snRNP und dem Spleißfaktor U2AF über kurze Introns (Intron-Erkennungsmodell) bzw. über kurze Exons (Exon-Erkennungsmodell) hinweg statt (Deutsch & Long 1999) (Abb. 4).



**Abb. 4: Intron- und Exon-Erkennungsmodelle** (modifiziert nach De Conti et al. 2013): Die schwarzen Rechtecke symbolisieren Exons einer prä-mRNA, die durch Introns (schwarze Linien) getrennt sind. **A:** Sind große Exons durch kleine Introns (<250 nt) separiert, findet die Spleißstellenerkennung durch die Interaktion zwischen den Spleißosomkomponenten U1-snRNP und U2AF über die Introns hinweg statt. **B:** Große Introns (>250 nt) und kleine Exons führen zu einer Spleißstellenkommunikation über die Exons hinweg.

Da die prä-mRNAs von Wirbeltieren im Durchschnitt 137 nt lange Intron-flankierte Exons (interne Exons) besitzen, die von durchschnittlich 3400 nt langen Introns separiert sind, findet bei höheren Eukaryoten überwiegend das Exon-Erkennungsmodell Anwendung. Aufgrund dieser Tatsache kann der hohe Anteil des alternativen Spleißereignisses  $E_{aus}$  in Säugetieren durch eine Störung im Exon-Erkennungsmodell begründet werden (Robberson et al. 1990; Berget 1995; Deutsch & Long 1999; Schellenberg et al. 2008).

Die Interaktion zwischen dem U1-snRNP und dem Spleißfaktor U2AF wird über *cis*-Elemente und *trans*-Faktoren, die Hauptregulatoren des prä-mRNA-Spleißens, vermittelt (Fairbrother & Chasin 2000; Sun & Chasin 2000; Roca et al. 2005).

### 3.4.2 Spleißregulatoren

Die für die Erfassung der Exon/Intron-Architektur der prä-mRNA erforderlichen kanonischen Sequenzelemente sind aufgrund ihrer Kürze und Degeneration nicht in der Lage, eine vollständige Erkennung der Spleißstellen zu gewährleisten. Zusätzliche spleißregulatorische Elemente sind notwendig, um geeignete 5'- und 3'-Spleißstellen zu erkennen. Dazu zählen neben den *cis*-

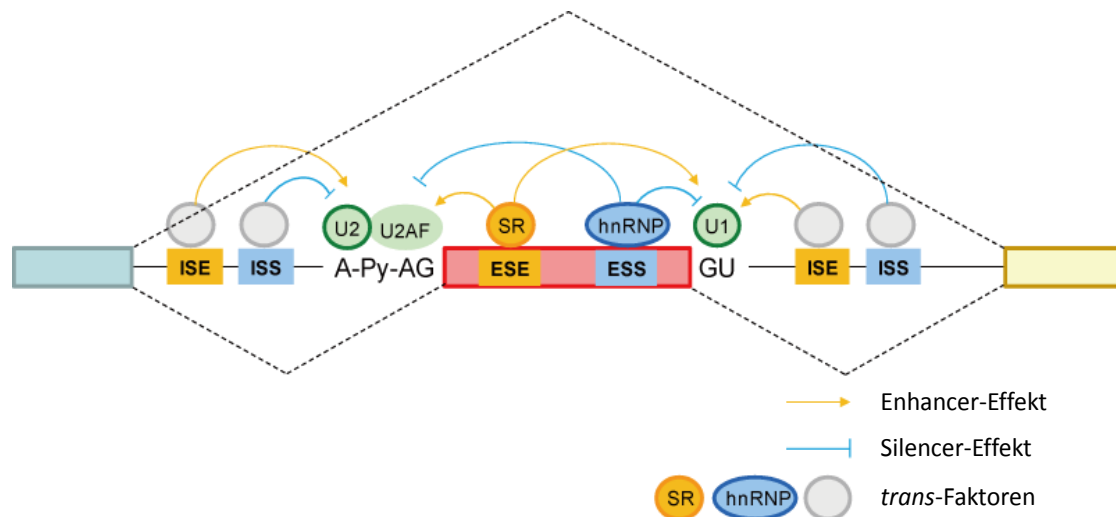
Elementen und *trans*-Faktoren, die maßgeblich an der Regulation einer Spleißstellenerkennung beteiligt sind, die prä-mRNA-Sekundärstruktur, die Histonmodifikationen, die Chromatinstruktur sowie die Interaktionen zwischen Spleißfaktoren und der RNA-Polymerase II (Cramer et al. 1997; Levine & Tjian 2003; Buratti & Baralle 2004; Hiller, Zhang, et al. 2007; Luco et al. 2011; De Conti et al. 2013). Des Weiteren wird durch den Abbau bestimmter Spleißvarianten (engl. *nonsense-mediated mRNA decay*, NMD) die Genexpression posttranskriptionell gesteuert (Hilleren & Parker 1999; Vilardell et al. 2000). Dieser Prozess der Transkriptkonzentrationsregulation wird als RUST (engl. *regulated unproductive splicing and translation*) bezeichnet und findet überwiegend Anwendung bei der Autoregulation von Spleißfaktoren (Jumaa & Nielsen 1997; Sureau et al. 2001; Hillman et al. 2004).

Im Mittelpunkt dieser Doktorarbeit stehen die *cis*-Elemente und *trans*-Faktoren. Zu den *cis*-Elementen zählen 6-8 nt lange prä-mRNA-Sequenzabschnitte (spleißregulatorische Motive, SRMs), die in Exons und Introns lokalisiert sind und die Spleißstellenerkennung in verstärkender oder hemmender Weise teils gewebe- oder zelltypspezifisch beeinflussen (Fedorov et al. 2001). Während in einer humanen prä-mRNA exonische Enhancer- (engl. *exonic splicing enhancer*, ESE) und Silencer-Sequenzen (engl. *exonic splicing silencer*, ESS) maximal 125 nt von einer Spleißstelle entfernt sind, beträgt der Abstand der intronischen Enhancer (engl. *intrinsic splicing enhancer*, ISE) und Silencer (engl. *intrinsic splicing silencer*, ISS) zur nächstgelegenen Exon-Intron-Grenze maximal 200 nt. Die auffällig geringe Anzahl an Einzelnukleotid-Polymorphismen (engl. *single nucleotide polymorphisms*, SNPs) in einer Region von 20 bp stromaufwärts und stromabwärts der Exon-Intron-Grenzen im humanen Genom weist auf eine direkte Nachbarschaft der SRMs zur Spleißstelle hin (Rosenow et al. 2002). Diese Hypothese wird von einer Studie unterstützt, die zeigt, dass die Variabilität der Kodonen am Exon-Ende abnimmt (Willie & Majewski 2004).

In den meisten Fällen werden *cis*-Elemente von *trans*-Faktoren gebunden, welche überwiegend nicht-snRNP Spleißfaktoren sind. Die *trans*-Faktoren lassen sich in zwei Hauptgruppen unterteilen. Die erste Gruppe bilden die SR- und SR-ähnlichen Proteine, die eine Arginin/Serin-reiche (RS) Domäne und eine zusätzliche RNA-Bindedomäne (engl. *RNA recognition motif*, RRM) aufweisen. RNA-Bindeproteine, die keine RS-Domäne besitzen, repräsentieren die zweite Gruppe. Hierzu zählen heterogene nukleäre Ribonukleoproteine (engl. *heterogeneous nuclear ribonucleoproteins*, hnRNPs), NOVA, FOX, MBNL und weitere RNA-Bindeproteine (Graveley 2000; Martinez-Contreras et al. 2007; Zhong et al. 2009).

Das komplexe Zusammenspiel von *cis*-Elementen und *trans*-Faktoren bildet eine spleißregulatorische Einheit, die in Wechselwirkung mit Spleißosomkomponenten tritt und eine Auswahl von alternativ gespleißten Exons zur Folge hat. Während die SR-Proteinfamilie überwiegend Enhancer-Sequenzen bindet und eine aktivierende Rolle im Spleißgeschehen einnimmt, hemmen hnRNPs die Spleißreaktion durch Bindung einer Silencer-Sequenz (Schaal &

Maniatis 1999; Black 2003). Dementsprechend vermitteln in Abbildung 5 Enhancer-Effekte einen Exon-Einschluss ( $E_{\text{ein}}$ ) durch die Rekrutierung der Spleißosomkomponenten U2AF, U1- und U2-snRNP an eine alternative 5'- und 3'-Spleißstelle. Im Gegensatz dazu verstärken Silencer den  $E_{\text{aus}}$  durch Inhibition der Spleißosom-Assemblierung an den benannten Spleißstellen.



**Abb. 5: Schematische Darstellung einer Spleißregulation** (modifiziert nach Z. Wang & Burge 2008): Die Rechtecke symbolisieren Exons einer prä-mRNA, die durch Introns (schwarze Linie) separiert sind. „GU“, „AG“, „A“ und „Py“ repräsentieren die kanonischen Sequenzelemente 5'- und 3'-Spleißstelle, Verzweigungspunkt sowie Polypyrimidintrakt, welche von den Spleißosomkomponenten U1-, U2-snRNP und vom Spleißfaktor U2AF erkannt werden. Das Spleißen wird durch die RNA-Protein- und Protein-Protein-Interaktionen der *cis*-Elemente (ESE, ESS, ISE, ISS) und der *trans*-Faktoren (SR-Proteine, hnRNPs und unbekannte Spleißfaktoren) reguliert, wobei ein Exon-Einschluss bzw. Exon-Ausschluss (gestrichelte Linien) generiert wird.

Bereits der Verlust eines einzigen *cis*-Elementes bzw. *trans*-Faktors kann zu biologisch bedeutsamen alternativen Spleißreaktionen führen. Beispielsweise verursacht eine Punktmutation (C>T) im Exon 7 an der Position +6 im humanen *SMN2*-Gen (engl. *survival of motor neuron 2*) eine Inaktivierung der Bindestelle des *trans*-Faktors SF2/ASF. Folglich wird bei der Mehrheit der Transkripte Exon 7 ausgeschlossen, wodurch ein C-terminal verkürztes, instabiles und vermutlich nicht funktionales Protein synthetisiert wird (Cartegni & Krainer 2002; Zhang et al. 2008). Der Mangel an funktionellen *SMN2*-Proteinen führt in Kombination mit dem Verlust des *SMN1*-Gens (engl. *survival of motor neuron 1*) zum Abbau von Motoneuronen in Stammhirn und Rückenmark und somit zu Spinaler Muskelatrophie (Wirth et al. 2006).

### 3.5 Bedeutung der 5'-Spleißstelle im Krankheitskontext

Neuesten Erkenntnissen zufolge unterliegen ca. 95 % der humanen Multi-Exon-Gene einem alternativen Spleißprozess (Pan et al. 2008). Dabei besitzen Fehlregulationen alternativer Spleißprozesse einen erheblichen Einfluss auf die Entstehung bzw. den Verlauf komplexer Erkrankungen. Bereits vor 20 Jahren konnten ca. 15 % der Punktmutationen, die innerhalb einer

Spleißstelle im humanen Genom lagen, mit einer genetischen Erkrankung in Zusammenhang gebracht werden (Krawczak et al. 1992).

Mutationen, die das Spleißgeschehen beeinflussen, lassen sich in zwei Gruppen unterteilen. Die erste Gruppe umfasst Mutationen, die in kanonischen Sequenzelementen lokalisiert sind. 9-10 % der bekannten genetischen Erkrankungen sind mit dieser Gruppe I assoziiert (Wang & Cooper 2007). Mutationen in exonischen und intronischen SRMs zählen zur Gruppe II. Laut aktuelleren Studien bewirken Mutationen wie SNPs, Mikro-Insertionen und -Deletionen ( $\leq 6$  bp) in exonischen SRMs (ESEs und ESSs) den Großteil an nicht-kanonischen Spleißereignissen (Faustino & Cooper 2003). In einigen Genen liegen bis zu 50 % der genspezifischen Mutationen in den exonischen SRMs. Auf Grundlage dieser Erkenntnisse wurde die Hypothese aufgestellt, dass über 60 % der krankheitsassoziierten Mutationen den nativen Spleißcode zerstören (López-Bigas et al. 2005).

Einen Schlüsselschritt des Spleißens stellt die Erkennung der 5'-Spleißstelle dar (Wang & Burge 2008). Bedeutende Vertreter der Auswirkung einer fehlerhaften 5'-Spleißstellen-erkennung sind die Erbkrankheiten FTDP-17 (Frontotemporale Demenz und Parkinsonismus des Chromosoms 17), Sarkoidose und Epilepsie-Anfälligkeit (Lee et al. 2001; Valentonyte et al. 2005; Heinzen et al. 2007).

Über 70 % der humanen Introns beinhalten multiple 5'-Spleißstellen, die während einer Spleißreaktion nicht erkannt werden und mit einer 98 %igen Wahrscheinlichkeit stromaufwärts ein vorzeitiges Translationsstoppkodon (engl. *premature termination codon*, PTC) aufweisen. Im Fall einer Aktivierung dieser latenten 5'-Spleißstellen wird eine vorzeitig terminierte mRNA prozessiert, die zur Entstehung verkürzter Proteine beiträgt, die wiederum eine Krankheitsrelevanz besitzen können (Miriami et al. 2002; Li et al. 2002; Miriami et al. 2004). Eine *in vitro* Aktivierung der latenten 5'-Spleißstellen im humanen Genom erfolgte bereits (Nevo et al. 2012).

Mutationen, im Speziellen Mutationen in exonischen und intronischen SRMs, die zur Erkennung der latenten 5'-Spleißstellen beitragen und zur Expression verkürzter Proteine führen können, müssen in Hinblick auf die molekulare Aufklärung diverser Erbkrankheiten spezifisch analysiert werden.

### 3.6 Programme zur Vorhersage von *cis*-Elementen und *trans*-Faktoren

*Cis*-Elemente und deren korrespondierende *trans*-Faktoren wurden bereits aufgrund ihrer biologischen Relevanz global in experimentellen und bioinformatischen Ansätzen analysiert. Die Ergebnisse wurden zur Entwicklung von computerbasierten Programmen genutzt, mit deren Hilfe SRMs und deren Bindepartner vorhergesagt werden können. In Tabelle 1 sind die bedeutendsten und aktuell verfügbaren Vorhersageprogramme aufgelistet.

Tab. 1: Vorhersageprogramme von *cis*-Elementen und *trans*-Faktoren

Programm	Datengrundlage
<i>ESEfinder 3.0</i>	Funktionelles <i>in vivo</i> oder <i>in vitro</i> SELEX (engl. <i>systematic evolution of ligands by exponential enrichment</i> , SELEX) (Cartegni et al. 2003)
<i>Splicing Rainbow</i>	Datensatz literaturbekannter regulatorischer Spleißmotive und krankheits-assoziiierter Mutationen (Stamm et al. 2006)
<i>SFmap</i>	Analyse bekannter Spleißfaktoren, unter Berücksichtigung des genetischen Umfeldes und der evolutionären Konservierung (Paz et al. 2010)
<i>RESCUE-ESE</i>	Statistische Analyse von Exon-Intron- und Spleißstellen-Kompositionen (Fairbrother et al. 2002)
<i>PESE</i>	Analyse des Konservierungsgrades von Wobble-Positionen und Sequenz-Motiv-Abundanzen zwischen exonischen Human-Maus-Orthologen (Goren et al. 2006)
<i>PESX</i>	Vergleich einer 8-mer Häufigkeitsverteilung in internen nicht-kodierenden Exons mit ungespleißten Pseudo-Exons und 5'-UTRs von Transkripten intronloser Gene (Zhang & Chasin 2004)

Ein Praxistest zur Bestimmung der Zuverlässigkeit der Vorhersageprogramme ergab, dass diese mit einer maximalen Genauigkeit von 17 % die SRMs im humanen Genom identifizieren (ElSharawy et al. 2009). In Anbetracht der unterschiedlichen Datengrundlage der Programme ist diese geringe Vorhersage-Zuverlässigkeit ein Indiz für die Komplexität des Spleißcodes, für dessen Entschlüsselung weitere Faktoren berücksichtigt werden müssen. Die Betrachtung weiterer Faktoren wie Locus- und Gewebespezifität lassen ebenfalls aktuellere, experimentelle Studien vermissen (Ke et al. 2011; Wen et al. 2013).

Um die Spleißaktivität von Mutationen und latenten 5'-Spleißstellen im humanen Genom vorhersagen zu können, sind die Ergebnisse der vorangegangenen Studien ungeeignet.

### 3.7 Zielstellung

Im Rahmen dieser Arbeit sollen experimentell spleißregulatorische Motive in der Nähe einer 5'-Spleißstelle in Zellkulturmodellen identifiziert werden. Mithilfe der identifizierten Motive sollen Vorhersagen spleißrelevanter Mutationen im humanen Genom ermöglicht werden.

Um dieses Ziel zu erreichen, muss zunächst ein neuartiges Spleißreportersystem entwickelt werden. Dazu soll ein Plasmid konstruiert werden, welches ein vollständiges humanes Gen mit flankierenden Fluoreszenzproteinsequenzen beinhaltet und die Expression dieses Konstrukts in eukaryotischen Zellen ermöglicht. Dabei soll das System so konzipiert sein, dass konstitutive und alternative Spleißereignisse fluoreszenzbasiert unterschieden werden können. Im Anschluss an die funktionelle Validierung des Spleißreportersystems sollen bioinformatisch spleißregulatorische Loci in 5'-Spleißstellennähe des Kandidatengens identifiziert und experimentell bestätigt werden. Durch die randomisierte Mutagenese dieser Bereiche und die tran-

---

siente Transfektion der generierten Mutagenesebanken in drei humane Zelllinien unterschiedlichen Gewebeursprungs, soll anschließend die relative Spleißaktivität für jedes einzelne Motiv berechnet werden. Dies soll durch die Tiefensequenzierung der aus den Zellen isolierten Transkripte erfolgen. Die auf diese Weise identifizierten Motive sollen entsprechend ihres Einflusses auf die Spleißaktivität in verstärkende (Enhancer) und hemmende (Silencer) Motive eingeteilt werden. Die Ergebnisse sollen durch fluoreszenzbasierte Analysen zufällig ausgewählter Motive validiert werden. Abschließend sollen *trans*-Faktoren, die die neuen, experimentell ermittelten spleißregulatorischen Motive binden, identifiziert werden. Hierfür soll das Analyseprogramm *SFmap* zur Anwendung kommen, das die identifizierten Motive bewertet und diese mit bereits charakterisierten RNA-Motiven, deren Bindepartner bekannt sind, abgleicht.

## 4 MATERIAL UND METHODEN

### 4.1 Chemikalien

Die verwendeten Chemikalien wurden, wenn nicht anders angegeben, von Carl Roth GmbH & Co. KG, Karlsruhe in analysenreiner Qualität bezogen.

Im Zuge der Doktorarbeit kam ausschließlich bidestilliertes Wasser zum Einsatz, welches ebenso wie die Puffer, Lösungen und Medien autoklaviert wurde.

### 4.2 Herstellung rekombinanter Plasmide

#### 4.2.1 Polymerase-Ketten-Reaktion (PCR)

Die Methode der Polymerase-Ketten-Reaktion (engl. *polymerase chain reaction*, PCR) wurde 1986 von Mullis et al. beschrieben und in einem Thermocycler (*T3000 Thermocycler* von Biometra, Göttingen) durchgeführt (Mullis et al. 1986).

Für das Primer-Design wurden die Programme *OligoCalc* (Kibbe 2007) und *Primer3 v.0.4.0* (Rozen & Skaletsky 2000) genutzt.

Die in den folgenden Abschnitten eingesetzten DNA-Polymerasen, dNTPs (Promega, Mannheim) und Primer (Microsynth, Balgach, Schweiz) kamen unter Einhaltung der Herstellerangaben zum Einsatz. Alle verwendeten Primer sind im Anhang I aufgelistet.

Die Standard-PCR dieser Doktorarbeit wurde unter Verwendung der *GoTaq<sup>®</sup> DNA Polymerase* (Promega, Mannheim), sofern nicht anders angegeben, wie folgt durchgeführt:

<b>Standard-PCR:</b>	Initiale Denaturierung:	95 °C, 2 min	28 Zyklen
	Denaturierungsphase:	95 °C, 30 s	
	Annealing:	56 °C, 30 s	
	Elongationsphase:	72 °C, 1 min pro kb	
	Finale Elongation:	72 °C, 5 min	
	Abkühlung:	4 °C, ∞	

- **Temperaturgradienten-PCR**

Zur Optimierung verschiedener PCR-Reaktionsparameter, z.B. Annealing-Temperatur, Reduzierung von Primer-Dimeren und unspezifischen Produkten, wurde für verschiedene Primer-Paare mithilfe eines Gradientencyclers (*MyCycler<sup>™</sup> Thermal Cycler v1.065* von Bio-Rad Laboratories, München) eine Temperaturgradienten-PCR durchgeführt. Hierbei wurden Annealing-Temperaturen zwischen 70 °C und 53 °C getestet.



- **Amplifizierung genomischer Fragmente**

Zunächst erfolgte eine Sequenzauswahl genomischer Fragmente mithilfe der Gendatenbank *UCSC Human Genome Browser* (<<http://genome.ucsc.edu>>, Version NCBI36/hg18). Die Datenbank enthält zudem Informationen über chromosomale Lage, Exon-Intron-Übergänge, Konservierungsgrad, gewebespezifische Expressionslevel und Spleißereignisse der möglichen Kandidatengene.

Zur Amplifizierung genomischer Fragmente wurde ein Touchdown-PCR-Programm verwendet. Die eingesetzten Primer beinhalteten zusätzlich eine Erkennungssequenz für eine Restriktionsendonuklease am 5'-Ende. Als Template wurde CEPH-DNA (*Centre d'Etude du Polymorphisme Humain*, CEPH) genutzt. Eine Amplifikation wurde mit dem enzymatischen Mix *Platinum® Taq DNA Polymerase High-Fidelity* (Invitrogen, Darmstadt), bestehend aus einer *Pfu*- und einer *Taq*-DNA-Polymerase, durchgeführt. Die *Pfu*-DNA-Polymerase gewährleistete aufgrund einer Proofreading-Aktivität eine geringe Fehlerrate.

<b>Touchdown-PCR:</b>	Initiale Denaturierung:	95 °C, 2 min	40 Zyklen
	Denaturierungsphase:	94 °C, 30 s	
	Annealing (Touchdown in 12 Schritten):		
		65-53 °C à 30 s ( $\Delta 1$ °C pro Zyklus, 12 Zyklen);	
		53 °C à 30 s (28 Zyklen)	
	Elongationsphase:	68 °C, 1 min pro kb	
	Finale Elongation:	72 °C, 5 min	
	Abkühlung:	4 °C, $\infty$	

PCR-Produkte mit einem 3'-Adenin-Überhang konnten nach Bedarf in einem pCR2.1-TOPO-Vektor subkloniert werden (Abschn. 4.2.5).

- **Ortsgerichtete Mutagenese**

Mutationen wie Insertionen, Deletionen und Basenaustausche wurden durch ortsgerichtete Mutagenesen, unter Verwendung des *QuickChange® Lightning Site-Directed Mutagenesis Kits* (Agilent Technologies, Böblingen), nach Herstellerangaben, umgesetzt. Das Design von ortsgerichteten Mutagenese-Primern erfolgte mit dem *QuickChange® Primer Design Tool* der Agilent-Technologies-Homepage.

- **Randomisierte Mutagenese**

Zur Herstellung einer randomisierten Mutagenesebank wurde die *Phusion High-Fidelity DNA Polymerase* (New England Biolabs, Frankfurt am Main), nach Herstellerangaben, eingesetzt. Die Polymerase ist speziell für die Generierung langer Amplikons (bis zu 20 kb) geeignet und

besitzt zudem aufgrund einer 3'-5'-Exonuklease-Proofreading-Aktivität eine extrem niedrige Fehlerrate ( $4,4 \times 10^{-7}$ ).

<b>Randomisierte Mutagenese:</b>	Initiale Denaturierung:	98 °C, 1 min	22 Zyklen
	Denaturierungsphase:	98 °C, 15 s	
	Annealing:	60,5 °C, 30 s	
	Elongationsphase:	72 °C, 30 s pro kb	
	Finale Elongation:	72 °C, 10 min	
	Abkühlung:	4 °C, ∞	

Aufgrund des Einsatzes 5'-phosphorylierter Primer konnten die gereinigten, Template-freien PCR-Produkte im Anschluss zirkularisiert werden (Abschn. 4.2.2-3 und 4.2.6-7).

#### 4.2.2 Agarose-Gelelektrophorese

##### 1x TBE-Puffer pH 8,0

Substanz	Molarität
Tris	89 mM
Borat	89 mM
EDTA-Natriumsalz	2 mM

##### 5x Ladepuffer pH 8,0

Substanz	Vol.% / Molarität
EDTA-Natriumsalz	0,1 M
Glycerin	40 % (v/v)
Bromphenolblau	0,25 % (v/v)

Zur Größenaufftrennung der verschiedenen DNA-Produkte fanden 1-2 %ige Agarosegele Verwendung. 1-2 % (w/v) Agarose (Biozym, Hessisch Oldendorf) wurde hierfür in TBE-Puffer gelöst. Vor dem Polymerisationsprozess wurden 0,001 % (w/v) Ethidiumbromid hinzugefügt. Die Einstellung der Nukleinsäureproben auf eine 1x Ladepufferkonzentration erfolgte vor dem Beladen. Die Gelelektrophorese wurde in 1x TBE-Puffer bei konstanten 110-150 V durchgeführt. Als Größenstandard wurden 5 µl *SmartLadder* (Eurogentec, Köln) mitgeführt. Nach einer Elektrophoresezeit von 60-80 min wurden die DNA-Produkte auf einem UV-Tisch mit Kamera (*Alphamager<sup>®</sup>EP* von Biozym, Hessisch Oldendorf) sichtbar gemacht und dokumentiert. Nach Bedarf konnten einzelne DNA-Fragmente auf dem UV-Tisch aus dem Agarosegel mithilfe eines Skalpell herausgeschnitten werden. Die anschließende DNA-Gelextraktion erfolgte wie in Abschnitt 4.2.3 beschrieben.

### 4.2.3 Aufreinigung von DNA-Fragmenten

DNA-Gelextraktionen und Aufreinigungen von PCR-, Restriktions- und Ligationsprodukten erfolgten nach dem Standardprotokoll des *Wizard® SV Gel and PCR Clean-Up Systems* (Promega, Mannheim). Das Elutionsvolumen betrug 25 µl.

### 4.2.4 Konzentrationsbestimmung von Nukleinsäuren

Die Konzentrationsbestimmung von Nukleinsäuren erfolgte durch spektrophotometrische Messung der optischen Dichte (OD) bei 230 nm (Absorptionsmaximum von Polysacchariden), 260 nm (Absorptionsmaximum von Nukleinsäurebasen) und 280 nm (Absorptionsmaximum von Proteinen) mittels Spektrophotometer (*NanoPhotometer™* von Implen, München). Das Lösungsmedium der Proben wurde als Referenz genutzt. Für die Reinheitskontrollen wurden aus den OD-Werten die Verhältnisse 260/230 und 260/280 gebildet:

$$\text{Reinheit} = \frac{A_{260}}{A_{230}} \geq 2,0$$

**Bestimmung der Reinheit eines Polysaccharid-Nukleinsäure-Gemisches:** Die Reinheit eines Polysaccharid-Nukleinsäure-Gemisches errechnet sich aus dem Quotienten der OD-Werte bei 260 nm und 230 nm. Eine Probe gilt als rein, d.h. frei von Polysacchariden, wenn der Quotient  $\geq 2,0$  ist.

$$\text{Reinheit} = \frac{A_{260}}{A_{280}} \geq 1,7$$

**Bestimmung der Reinheit eines Protein-Nukleinsäure-Gemisches:** Die Reinheit eines Protein-Nukleinsäure-Gemisches errechnet sich aus dem Quotienten der OD-Werte bei 260 nm und 280 nm. Eine Probe gilt als rein, d.h. frei von Proteinen, wenn der Quotient  $\geq 1,7$  ist.

### 4.2.5 Kommerzielle Vektorsysteme

- **Topo-TA-Klonierung**

Die Subklonierung von PCR-Produkten, die mit einer *Taq*-DNA-Polymerase amplifiziert wurden, erfolgte mithilfe des *TOPO® TA Cloning® Kits (with pCR®2.1-TOPO Vector)* (Invitrogen, Darmstadt) nach Standardprotokoll des Herstellers. Der pCR2.1-TOPO-Vektor besitzt eine Topoisomerase I-vermittelte Ligationsfunktion. Der *lac*-Promoter induziert die Expression des *lacZα*-Fragmentes, welches die ersten 146 Aminosäuren der  $\beta$ -Galaktosidase kodiert und für eine Blau-Weiß-Selektion genutzt werden kann. Zusätzlich vermittelt der Vektor eine Kanamycin- und Ampicillinresistenz. Subklonierte DNA-Fragmente wurden mittels Sanger-Methode überprüft (Abschn. 4.2.10) und konnten nach Bedarf in das Zielploid transferiert werden (Abschn. 4.2.6-7).

- **Expressionsvektoren**

Die Basis der konstruierten Spleißreporter bildeten die Expressionsvektoren pAcGFP1-N1 und pDsRed-Express-N1 (Clontech, Heidelberg) (Abb. 6).

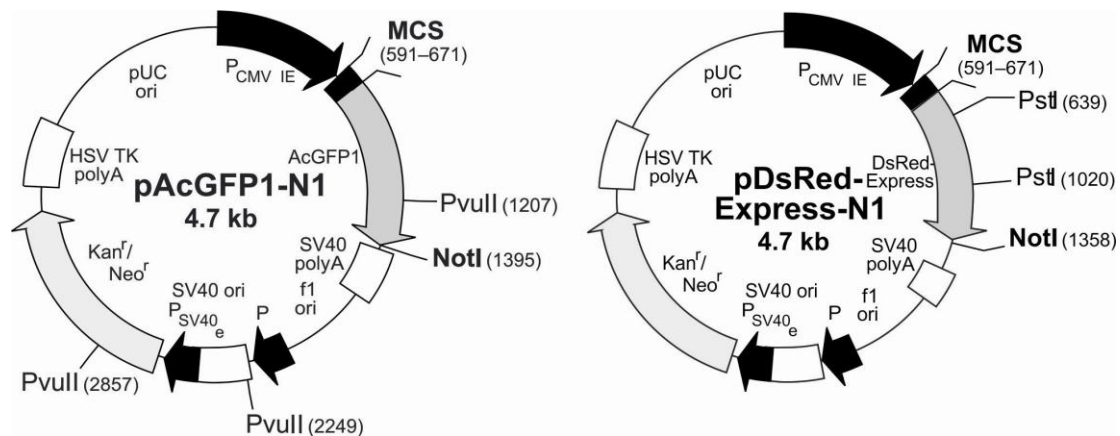


Abb. 6: Aufbau der Expressionsvektoren pAcGFP1-N1 und pDsRed-Express-N1 (nach Clontech 2005)

In diesem Vektorsystem werden die in die multiple Klonierungsstelle (engl. *multiple cloning site*, MCS) zu inserierenden genomischen Fragmente von den kodierenden Sequenzen der fluoreszierenden Proteine AcGFP1 (grün fluoreszierendes Protein (GFP) aus *Aequorea coerulescens*) bzw. DsRed-Express (rot fluoreszierende Proteinvariante (RFP) aus *Discosoma sp.*) in 3'-Richtung flankiert. Eine funktionelle Expression des rekombinanten Fusionsproteins in eukaryotischen Zellen setzt dabei den identischen Leserahmen der Insert-DNA und der Fluorophore voraus.

#### 4.2.6 DNA-Restriktion

Das Auffinden von Restriktionsschnittstellen in DNA-Sequenzen erfolgte mithilfe des Programms *NEBcutter V2.0* der Homepage von *New England Biolabs*.

Vektoren und PCR-Produkte wurden mit Restriktionsendonukleasen unter dem Einsatz geeigneter Puffer hydrolysiert. Folgende Restriktionsenzyme kamen zum Einsatz: BamHI, DpnI, MluI, NheI-HF, Sall-HF und XhoI (New England Biolabs, Frankfurt am Main). Die Verwendung der Enzyme fand gemäß der Herstellerangaben statt. Analytische Restriktionen (bis zu 300 ng DNA) wurden im 10 µl Maßstab durchgeführt und 30 min bei 37 °C inkubiert. Für einen präparativen Verdau (bis zu 2000 ng DNA), mit einem Gesamt-Reaktionsvolumen von 50 µl, betrug die Reaktionszeit bei 37 °C zwei Stunden. Nach dem Restriktionsverdau wurde die linearisierte DNA dokumentiert (Abschn. 4.2.2) und bei Bedarf aufgereinigt (Abschn. 4.2.3).

#### 4.2.7 Ligation und Zirkularisierung von linearer DNA

Für die Ligation zweier doppelsträngiger DNA-Moleküle kam das Enzym T4-DNA-Ligase (New England Biolabs, Frankfurt am Main), nach Herstellerangaben, zum Einsatz. Vektor- und Insert-DNA wurden in einem Molaritätsverhältnis von 1:3, bei einer Gesamtmenge an Vektor-DNA von 100 ng, eingesetzt. Nach Ablauf der Reaktionszeit erfolgte eine direkte Transformation des Ligationsansatzes in *E.coli*-Zellen (Abschn. 4.2.8).

Eine Zirkularisierung 5'-phosphorylierter PCR-Produkte erfolgte ebenfalls mit dem Enzym T4-DNA-Ligase, ausgeführt in einem 250 µl Ansatz nach Herstellerangaben. Im Anschluss wurden die aufgereinigten Ligationsansätze (Abschn. 4.2.3) mittels Gefriertrocknung (Gefriertrocknungsanlage *ALPHA 1-2 LD plus* von Martin Christ GmbH, Osterode) auf eine Konzentration von 100 µg/µl eingestellt.

#### 4.2.8 Transformation von Plasmid-DNA in kompetente *E.coli*-Zellen

##### LB-Agar pH 7,5

Substanz	Menge
Bacto-Hefeextrakt	5 g/l
Bacto-Trypton	10 g/l
NaCl	5 g/l
Bacto-Agar	15 g/l

Unter der Ausführung des Standardprotokolls *One Shot® Chemical Transformation* (Invitrogen, Darmstadt) wurde Plasmid-DNA in *Top10*-Zellen (*One Shot® Chemically Competent E.coli*) (Invitrogen, Darmstadt) transformiert. Die Transformation ortsgerichteter Mutagenesen erfolgte nach Herstellerangaben in *XL10-Gold Ultracompetente Zellen* (Agilent Technologies, Böblingen) (Abschn. 4.2.1). Ein Teil des Transformationsansatzes wurde anschließend auf LB-Agar ausplattiert. Die mit einem pCR2.1-TOPO-Vektor (Abschn. 4.2.5) transformierten Bakterien wurden im Zuge einer Blau-Weiß-Selektion auf 30 g/l IPTG- und 20 g/l X-Gal-haltigen LB-Agar-Platten ausgestrichen. Aufgrund der Zugabe von 100 µg/ml Ampicillin bzw. 25 µg/ml Kanamycin in den LB-Agar, fand eine Selektion transformierter Bakterien statt. Die Kultivierung erfolgte über Nacht bei 37 °C.

#### 4.2.9 Isolierung von Plasmid-DNA aus *E.coli*

##### LB-Medium pH 7,5

Substanz	Menge
Bacto-Hefeextrakt	5 g/l
Bacto-Trypton	10 g/l
NaCl	5 g/l

Für eine Plasmidpräparation im kleinen Maßstab (Miniprep-Verfahren) erfolgte die Überführung transformierter Bakterien in 5 ml LB-Medium. Wurde eine sehr große Menge an Plasmid-DNA benötigt, fand eine Weiterkultivierung der plasmidtragenden *E.coli*-Zellen in 200 ml LB-Medium statt (Maxiprep-Verfahren). Um den Selektionsdruck aufrecht zu erhalten, wurde dem LB-Medium 100 µg/ml Ampicillin bzw. 25 µg/ml Kanamycin zugegeben. Es folgte eine Übernachtin-

kubation bei 37 °C. Die Isolierung von Plasmid-DNA im Miniprep-Maßstab erfolgte mit dem *Wizard® Plus SV Minipreps DNA Purification System* (Promega, Mannheim). Plasmid-DNA-Isolierungen im Maxiprep-Verfahren wurden mittels *QIAfilter™ Plasmid Maxi Kit* (Qiagen, Hilden) durchgeführt. Die Umsetzung der Standardprotokolle fand gemäß der Herstellerangaben statt.

#### 4.2.10 Sanger-Sequenzierung

Die DNA-Sequenzierung wurde, falls nicht anders angegeben, nach der Sanger-Kettenabbruchmethode durchgeführt (Sanger & Coulson 1975; Sanger et al. 1977). Für den Reaktionsansatz kam das *BigDye Terminator v3.1 Cycle Sequencing Kit* (Applied Biosystems, Darmstadt), nach Herstellerangaben, zum Einsatz. Folgendes Sequenzierprogramm wurde ausgeführt:

<b>Sequenzierprogramm:</b>	Initiale Denaturierung:	96 °C, 1 min	30 Zyklen
	Denaturierungsphase:	96 °C, 10 s	
	Annealing:	50 °C, 5 s	
	Elongationsphase:	60 °C, 4 min	
	Abkühlung:	4 °C, ∞	

Im Anschluss wurden die DNA-Produkte im Institut für Klinische Molekularbiologie der Christian-Albrechts-Universität zu Kiel mit *DyeEx®*-Säulen (Qiagen, Hilden) aufgereinigt und mithilfe eines Kapillarsequenzierers (*ABI PRISM® 3700 DNA Analyzer* von Applied Biosystems, Foster City, Kalifornien, USA) sequenziert.

Die Auswertung der Sequenzdaten erfolgte mit dem Programm *Sequencher 4.5* (Gene Codes Corporation).

### 4.3 Zellkultur

Die in dieser Arbeit verwendeten HeLa-Zellen (humane Gebärmutterhalskrebszellen, ACC 57) stammten von der Deutschen Sammlung für Mikroorganismen und Zellkulturen GmbH, Braunschweig (DSMZ). Die Zelllinie SH-SY5Y (humane Neuroblastomzellen) wurde von der Abteilung für Genomforschung des Fritz-Lipmann-Institutes e.V. Jena bereitgestellt und die Zelllinie HuH7 (hepatozelluläre Karzinomzellen) wurde von der Arbeitsgruppe für Molekulare Gastroenterologie der I. Medizinischen Klinik des Universitätsklinikums Schleswig-Holstein bezogen.

#### 4.3.1 Kultivierung von eukaryotischen Zellen

Alle Arbeitsschritte der Kultivierung und Behandlung von Zellen fanden unter sterilen Bedingungen statt. Die Anzucht und Kultivierung der adhärenenten Zelllinien erfolgte in kollagen-

beschichteten T75-Kulturflaschen (BD Falcon, Heidelberg) bei 37 °C, gesättigter Luftfeuchte und einem CO<sub>2</sub>-Gehalt von 5 %. Die Zelllinien SH-SY5Y und HuH7 wurden in DMEM High Glucose (4,5 g/l)-Medium (PAA Laboratories, Pasching) und HeLa-Zellen in RPMI 1640-Medium (PAA Laboratories, Pasching) kultiviert. Die Versorgung der Zellen mit Wachstumsfaktoren wurde durch die Zugabe von 10 % (v/v) fetalem Kälberserum (FCS) (Biochrom, Berlin) in das Kulturmedium sichergestellt. Nach einer visuellen Einschätzung der Zelldichte unter einem Inversmikroskop (*Olympus CK2* von Olympus Deutschland, Hamburg) erfolgte ab einer 80 %igen Konfluenz das Passagieren der Zellen. Hierfür wurden die Zellen mit PBS (PAA Laboratories, Pasching) gewaschen und dann mit 1x Trypsin-EDTA-Lösung (Gibco, Eggenstein) für fünf Minuten bei 37 °C inkubiert. Die Enzymbehandlung wurde durch Zugabe von Kulturmedium beendet, die Zellsuspensionen in 50-ml-Röhrchen (BD Falcon, Heidelberg) überführt und pelletiert (5 min, Raumtemperatur und 500×g) (Zentrifuge 5804 R von Eppendorf, Köln). Nach Resuspendieren der sedimentierten Zellen in 1-2 ml Medium erfolgte die Weiterkultivierung von 3-5×10<sup>5</sup> Zellen. Einmal pro Monat wurden die Zellen auf Kontamination mit Mykoplasmen untersucht. Hierfür kam das *PCR Mycoplasma Test Kit I/C* (Promokine, Heidelberg) zur Anwendung, dessen Standardprotokoll exakt eingehalten wurde.

#### 4.3.2 Zellernte und Zellzahlbestimmung

Zum Ernten der Zellen fand ein Austausch des Kulturmediums gegen 4 °C kaltes PBS statt. Dieser Waschschrift wurde mindestens zweimal wiederholt. Mit einem Gummischaber wurden die Zellen in PBS von der Kulturschale abgelöst und die Zellzahl mittels Neubauer-Zählkammer (VWR International, Darmstadt) bestimmt. Eine 1:2 Verdünnung, bestehend aus Zellsuspension und Trypanblau (0,1 %ig), das von toten Zellen nicht aus dem Zytosol transportiert werden kann, wurde in eine Neubauer-Zählkammer pipettiert. Die Berechnung der Lebendzellzahl für je 1 ml Zellsuspension erfolgte mit der Formel:

$$Z \times V \times 10^4 = \text{Zellzahl [ml}^{-1}\text{]}$$

**Zellzahlbestimmung mittels Neubauer-Zählkammer:** Die Zellzahl je 1 ml Zellsuspension berechnet sich aus dem Produkt der durchschnittlichen Zellzahl je Großquadrat (Z) und dem Verdünnungsfaktor (V), bei einem Kammervolumen von 10 µl.

Im Rahmen von FACS-Analysen (Abschn. 4.5.1) erfolgte die Überführung der Zellsuspensionen in 5-ml-FACS-Reaktionsgefäße (Sarstedt, Nümbrecht). Waren DNA- bzw. RNA-Analysen vorgesehen, wurden die Zellen in 1,5-ml-Reaktionsgefäße transferiert und zentrifugiert (5 min, 4 °C und 300×g) (Zentrifuge 5804 R von Eppendorf, Köln). Die gewonnen Zellpellets kamen für Gesamt-DNA- bzw. Gesamt-RNA-Isolierungen zum Einsatz (Abschn. 4.4.1 und 4.4.3) oder wurden bis zur weiteren Verwendung in flüssigem Stickstoff schockgefroren und bei -80 °C gelagert.

### 4.3.3 Transfektion von eukaryotischen Zellen

Für die Transfektionsexperimente wurden Zellsuspensionen definierter Zellzahlen hergestellt, in das versuchsabhängige Kulturgefäß ausgesät und über Nacht kultiviert (Abschn. 4.3.1-2). Nach 24 h hatten die Zellen eine 60-80 %ige Konfluenz und konnten für die Transfektionsexperimente eingesetzt werden. Für die Experimente wurden nur Zellen bis zur maximal 25. Passage verwendet.

Für eine transiente Transfektion der adhärenenten Zellen mit Plasmid-DNA kam das *Effectene* Reagenz (Qiagen, Hilden) zum Einsatz. Die biochemische Grundlage dieses Transfektionsreagenzes bildet ein Enhancer, mit dessen Hilfe DNA in Mizellen eines nicht-liposomalen Lipids in die Zellen transferiert wird. Eine hohe Transfektionseffizienz, bei gleichzeitig geringer Toxizität, wird in normalem, serumhaltigen Medium erzielt.

In dieser Arbeit fanden die Transfektionsversuche nach Standardprotokoll des Herstellers statt. Lediglich zur Transfektion der ineffizient transfizierbaren Zelllinie SH-SY5Y wurde die 1,5x Menge an empfohlener Effectene-Reagenz verwendet. Für Kotransfektionen wurden äquimolare Plasmidmengen eingesetzt, wobei die protokollbasierte Gesamt-DNA-Menge nicht überschritten wurde. Nach 48 h Inkubation erfolgten Ernte und Analysen.

Die Transfektionseffizienz wurde mittels FACS bestimmt (Abschn. 4.5.1).

## 4.4 DNA- und RNA-Analysen

### 4.4.1 Isolierung von Gesamt-DNA aus eukaryotischen Zellen

Zur Isolierung von Gesamt-DNA aus bis zu  $2 \times 10^6$  eukaryotischen Zellen fand die Verwendung des *QIAamp DNA Mini Kits* (Qiagen, Hilden), ausgeführt nach Herstellerangaben, statt. Die gewonnene Gesamt-DNA wurde im Anschluss zur Bestimmung von Gen- und Plasmidkopienzahlen eingesetzt (Abschn. 4.4.2).

### 4.4.2 Quantitative Real-Time-PCR (qRT-PCR)

Die Methode der quantitativen Real-Time-PCR (qRT-PCR) ermöglicht neben der herkömmlichen Target-Amplifizierung eine zusätzliche Quantifizierung der Amplikons. Die Quantifizierung wird mithilfe von Fluoreszenzmessungen durchgeführt. Die Fluoreszenz nimmt proportional zur PCR-Produktmenge zu. Eine theoretische Größe, der  $C_t$ -Wert (engl. *cycle threshold*), beschreibt dabei den Teil einer PCR-Zykluskurve, in dem die Fluoreszenz erstmals exponentiell über den Hintergrundwert ansteigt.

Im Zuge dieser Arbeit wurde die qRT-PCR zur Bestimmung von Gen- und Plasmidkopienzahlen eingesetzt (Joshi et al. 2008). Die Berechnung dieser Kopienzahl pro transient transfizierter



eukaryotischer Zelle erfolgte durch das Erstellen einer Plasmidstandardkurve aus definierten Plasmidkopienzahlen und deren zugehörigen  $C_t$ -Werte.

Zum Nachweis der Existenz eines diploiden Chromosomensatzes musste ein beliebiges, humanes Genfragment in einem pCR2.1-TOPO-Vektor kloniert (Abschn. 4.2.1-5 und 4.2.8-10) und als endogene Kontrolle verwendet werden. Kloniert wurde ein 149 bp Ausschnitt des C5-Gens (fünfte Komponente des Komplements).

Die qRT-PCR wurde mittels Real-Time-Cycler (*MyiQ™ Single-Color Real-Time PCR Detection System* von Bio-Rad Laboratories, München) und *Maxima™ SYBR Green/ROX qPCR Master Mix* (Fermentas, St. Leon-Rot), nach Herstellerangaben, ausgeführt. Die verwendeten Primer (Microsynth, Balgach, Schweiz) sind im Anhang I aufgelistet. Das qRT-PCR-Programm wurde wie folgt durchgeführt:

<b>qRT-PCR:</b>	Initiale Denaturierung:	95 °C, 10 min	40 Zyklen
	Denaturierungsphase:	95 °C, 20 s	
	Annealing:	58 °C, 20 s	
	Elongationsphase:	72 °C, 1 min pro kb	
	Abkühlung:	4 °C, ∞	

Die Berechnung der Gen- und Plasmidkopienzahlen erfolgte mit dem Programm *Microsoft Excel 2007* (Microsoft).

#### 4.4.3 Isolierung von Gesamt-RNA aus eukaryotischen Zellen

Zur Isolierung von Gesamt-RNA aus eukaryotischen Zellen kam das *RNeasy® Mini Kit* (Qiagen, Hilden) zum Einsatz. In Abhängigkeit der einzelnen Versuche wurden bis zu  $2 \times 10^6$  Zellen geerntet (Abschn. 4.3.2) und nach Standardprotokoll des Herstellers isoliert. DNA-Kontaminationen in den RNA-Proben wurden durch einen doppelten DNaseI-Verdau, à 30 min Inkubationszeit bei Raumtemperatur, eliminiert. Hierfür wurden 70 µl RDD-Puffer (Qiagen, Hilden) und 10 µl DNaseI (RNase-frei) (Qiagen, Hilden) auf die *RNeasy Mini Spin Column* gegeben. Das Elutionsvolumen betrug 50 µl.

Zur Überprüfung der RNA-Proben auf Kontaminationen mit genomischer DNA bzw. auf Verunreinigung mit Plasmid-DNA, im Rahmen von Transfektionsexperimenten, wurde ein genomisches Fragment des *GAPDH*-Gens (Glycerinaldehyd-3-Phosphat-Dehydrogenase) bzw. des Plasmids aus 100-200 ng Gesamt-RNA, unter Verwendung des Standard-PCR-Programms mit 32 Zyklen, amplifiziert (Abschn. 4.2.1). Als Positivkontrolle diente HeLa-Zellen-cDNA bzw. Plasmid-DNA. Kontaminierte Proben wurden nochmals auf eine *RNeasy*-Säule gegeben und mit DNaseI verdaut. Im Anschluss wurden die RNA-Proben direkt zur cDNA-Synthese eingesetzt (Abschn. 4.4.4) oder bis zur weiteren Verwendung bei -80 °C gelagert.

#### 4.4.4 cDNA-Synthese und Reverse-Transkriptase-PCR (RT-PCR)

Um Transkripte amplifizieren zu können, wurde die aus eukaryotischen Zellen isolierte Gesamt-RNA (Abschn. 4.4.3) in komplementäre cDNA (engl. *complementary* DNA) umgeschrieben. Hierfür kam das Standardprotokoll des *Advantage<sup>®</sup> RT-for-PCR Kits* (Clontech, Heidelberg), unter Verwendung von randomisierten Hexamer-Primern, nach Herstellerangaben zum Einsatz. Zur Überprüfung einer erfolgreichen cDNA-Synthese und zum Konzentrationsabgleich der cDNA-Proben, wurde mittels Standard-PCR (Abschn. 4.2.1) zunächst das Transkript des Referenzgens *GAPDH* amplifiziert. Als Template-DNA für RT-PCRs dienten im Anschluss 2 µl der cDNA. Sofern nicht anders vermerkt, erfolgte die Ausführung des RT-PCR-Programms nach Standard-PCR-Protokoll mit 35 Zyklen (Abschn. 4.2.1).

#### 4.4.5 Kapillarelektrophorese-laserinduzierter Fluoreszenz (CE-LIF)

Die Kapillarelektrophorese-laserinduzierter Fluoreszenz (engl. *capillary electrophoresis with laser-induced fluorescence detection*, CE-LIF) stellt eine Variante der Elektrophorese dar, bei der die exakte Länge verschiedener DNA-Fragmente bestimmt werden kann. Die mit Fluoreszenzfarbstoffen markierten DNA-Proben werden, in Abhängigkeit ihrer Längenausdehnung, mittels Kapillarsequenzierer getrennt. Entstehende Lichtsignale werden in digitale Signale umgewandelt und als Elektropherogramm dargestellt. Im Elektropherogramm wird jedes detektierte DNA-Fragment als Peak abgebildet, dessen Höhe der Intensität des Lichtsignals entspricht. Die resultierende Peak-Fläche ist ein Maß für die Quantität eines DNA-Fragmentes. Die Länge des detektierten DNA-Fragmentes wird über die Messung des Abstandes zwischen dem Anfangspunkt des Elektropherogramms und dessen Peak berechnet. Zum Vergleich wird ein DNA-Standard aufgetragen, der mit einem anderen Fluoreszenzfarbstoff markiert ist, sodass die Peaks der Probe und des Standards voneinander unterschieden werden können (Imle 2005; Akbari et al. 2008; Schindler et al. 2009).

Zur Quantifizierung von Spleißvarianten wurde eine RT-PCR (Abschn. 4.4.4) mit 40 Zyklen und unter Verwendung eines 6-Carboxyfluoreszein-(6-FAM)-markierten Primers (Anhang I) durchgeführt. Ausgehend vom Analyse-Exon erfolgte die Primer-Hybridisierung in den flankierenden Exons. Im Anschluss wurden die fluoreszierenden RT-PCR-Produkte mit Wasser 1:40 verdünnt. 1 µl der Verdünnung wurde mit 10 µl Formamid ( $\geq 99,5\%$ ) und 0,5 µl des Standards *GeneScan GS500LIZ* (Applied Biosystems, Darmstadt) versetzt. Formamid verhinderte die Bildung von Sekundärstrukturen der PCR-Produkte und der Standard gewährleistete einen Fragmentlängenvergleich bis zu 500 bp. Es folgte eine dreiminütige Denaturierung bei 94 °C. Mithilfe eines Kapillarsequenzierers (*ABI 3730xl DNA Analyzer* von Applied Biosystems, Foster City, Kalifornien, USA) in der Abteilung für Genomforschung des Fritz-Lipmann-Institutes e.V. Jena wurden die einzelsträngigen RT-PCR-Produkte der Größe nach aufgetrennt.

Die Fluoreszenzeinheiten der den DNA-Fragmenten korrespondierenden Peak-Flächen wurden mittels *GeneMapper 4.0 Software* (Applied Biosystems) notiert. Die Quantifizierung der E<sub>aus</sub>- und E<sub>ein</sub>-Transkriptvarianten desselben Spleißreporters wurde durch die Bildung der Peak-Flächenverhältnisse mithilfe des Programms *Microsoft Excel 2007* (Microsoft) berechnet. Der Transkriptanteil der E<sub>ein</sub>-Isoform errechnet sich wie folgt:

$$X_{\text{Ein}} = \frac{PA_{\text{Ein}}}{PA_{\text{Ein}} + PA_{\text{Eaus}}}$$

**Quantifizierung des E<sub>ein</sub>-Transkriptanteils:** Der E<sub>ein</sub>-Transkriptanteil ( $X_{\text{Ein}}$ ) wird über das Peak-Flächenverhältnis der E<sub>ein</sub>- ( $PA_{\text{Ein}}$ ) und E<sub>aus</sub>-Isoformen ( $PA_{\text{Eaus}}$ ) bestimmt.

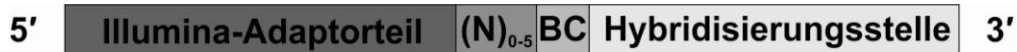
Im Zuge jedes CE-LIF-Experiments wurden drei unabhängige biologische Replikate angefertigt. Für die Signifikanzprüfung unabhängiger Proben mit beliebig vielen Messgrößen wurde ein zweiseitiger t-Test nach Student angewandt.

#### 4.4.6 Illumina-Tiefensequenzierung

Zur Tiefensequenzierung von 200-300 bp großer Amplikons der randomisierten Mutagenesebanken und Transkript-Pools kam die Next-Generation-Sequencing-Technologie *Illumina HiSeq 2000* zur Anwendung (Mardis 2008). Die Gesamtleselänge betrug 76 nt. Die verwendeten Forward- und Reverse-Primer zur Amplifikation der zu sequenzierenden Abschnitte beinhalteten stromaufwärts einen Illumina-Adaptorteil, der die Hybridisierung der Illumina-spezifischen Sequenzier-Primer ermöglichte. Dieser Illumina-Adaptor, mit einer Länge von bis zu 58 nt, musste über zwei PCRs synthetisiert werden. Zum Einsatz kam die bereits in Abschnitt 4.2.1 beschriebene *Phusion High-Fidelity DNA Polymerase*. Es wurden jeweils 35 PCR-Zyklen mit 56,4 °C Annealing-Temperatur ausgeführt. Im Anschluss wurden die PCR-Produkte von der Abteilung für Genomforschung des Fritz-Lipmann-Institutes e.V. Jena mit magnetischen Beads, unter Verwendung des *Agencourt® AMPure® XP 60mL Kits* (Beckman Coulter GmbH, Krefeld), aufgereinigt und mit dem *Illumina Genome Analyzer IIx* (Illumina, Inc., San Diego, Kalifornien, USA) sequenziert.

Für diese Doktorarbeit muss eine Besonderheit der Illumina-Sequenzierung hervorgehoben werden. Bei Amplikonsequenzierungen, die im Primer-Bereich homogene Sequenzgemische bilden, ist die Fluoreszenzaufnahme gestört und eine Sequenzanalyse nicht möglich. Aus diesem Grund wurden für die Amplifikation der zu sequenzierenden Region Forward-Primer verwendet, die stromabwärts des Illumina-Adaptorteils 0-5 randomisierte Basen besaßen. Diese Phasenverschiebung stellte eine Fluoreszenzaufnahme sicher. Somit ergaben sich sechs heterogene Forward-Primer-Stocks für eine Illumina-Sequenzierung. Des Weiteren waren Barcodes, bestehend aus drei Basen, stromabwärts der 0-5 randomisierten Basen im Forward-Primer integriert. Diese ermöglichten eine parallele Sequenzierung unterschiedlicher Experimente. In

Abbildung 7 ist der Forward-Primer zur Amplifikation der sequenzierten Abschnitte schematisch dargestellt.



**Abb. 7: Schematische Darstellung des Forward-Primers zur Amplifikation des Sequenzabschnitts der Tiefensequenzierung:** Der dargestellte Forward-Primer besteht aus einem Illumina-Adaptorteil, 0-5 Phasenverschiebungsbasen ((N)<sub>0-5</sub>), Barcode (BC) und einer Hybridisierungsstelle, angegeben in 5'-3'-Richtung.

Die von der Illumina-Sequenziersoftware *CASAVA* generierten 76 nt langen Sequenzfragmente (engl. *reads*) (FASTQ-Daten) wurden definierten Referenzsequenzen (Anhang II), in Abhängigkeit ihrer Homologie, zugeordnet. Der Sequenzabgleich wurde von der bioinformatischen Abteilung unserer Arbeitsgruppe unter Verwendung von *JavaScripten* ausgeführt.

Da sich proportional mit zunehmender Read-Länge die Illumina-Basenzuordnungsqualität verringert, wurden ausschließlich die ersten 65 Basenpositionen der 76 nt langen Reads ausgewertet.

Der Sequenzabgleich vollzog sich in zwei Arbeitsschritten. Zunächst wurden die Phasenverschiebungsbasen, die Barcodes und die Hybridisierungsstelle der Reads identifiziert. Insgesamt wurden eine Fehlbasenpaarung (engl. *mismatch*) in den Sequenzabschnitten der Phasenverschiebungsbasen und der Barcodes und zwei Insertionen bzw. Deletionen (Indels) in der Hybridisierungsstelle zugelassen. Während des zweiten Abgleichs mussten die ersten fünf benachbarten Basen der Mutationskassette exakt mit der Referenzsequenz übereinstimmen, lediglich ein Indel bzw. zwei Fehlbasenpaarungen wurden in den anschließenden 16 nt, stromaufwärts und stromabwärts, akzeptiert. Auf diesem Weg konnte für jedes erwartete Motiv eine spezifische Read-Anzahl generiert werden.

## 4.5 Proteinanalysen

### 4.5.1 Fluoreszenzaktivierte Zellsortierung (FACS)

Mithilfe der fluoreszenzaktivierten Zellsortierung (engl. *fluorescence activated cell sorting*, FACS) können eukaryotische Zellen in Suspension auf der Grundlage von Streulicht- und Fluoreszenzeigenschaften nach ihrer Größe, Granularität und Fluoreszenz beurteilt werden (Hulett et al. 1969). Die Durchflußzytometrie ermöglicht dabei eine hohe Analysegeschwindigkeit und Empfindlichkeit. Die Fluoreszenz der Zellen ist abhängig von dem Grad ihrer Eigenfluoreszenz und dem Expressionslevel der fluoreszierenden Proteine.

Die Anregung der Reportermoleküle in transfizierten Zellen (Abschn. 4.3.2-3) erfolgte im Durchflußzytometer (*BD FACSCalibur<sup>TM</sup>* von BD Biosciences, Heidelberg) mit einem blauen Argonlaser bei 488 nm. Der parallele Einsatz von *d2EGFP-1-* (488-507 nm) und *DsRed-Express-DR-* (557-579 nm) Filtern ermöglichte eine gleichzeitige Detektion der Fluorophore

AcGFP1 und DsRed-Express in einem Ansatz. Es wurden standardmäßig 10.000 Zellen pro Analyse gemessen.

In dieser Arbeit wurde das FACS zur Bestimmung von Transfektions- und Spleißeffizienzen eingesetzt. Transfektionseffizienzen errechneten sich aus dem Verhältnis der Fluoreszenzintensität transfizierter Zellen zu der Fluoreszenzintensität untransfizierter Zellen (Autofluoreszenz). Die Quantifizierung von  $E_{\text{aus}}$ - und  $E_{\text{ein}}$ -Proteinvarianten desselben Spleißreporters erfolgte durch die Bildung der GFP/RFP-Fluoreszenzintensitätsverhältnisse, mithilfe des Programms *Microsoft Excel 2007* (Microsoft). Der Proteinanteil der  $E_{\text{ein}}$ -Isoform errechnet sich wie folgt:

$$X_{\text{Ein}} = \frac{\text{GFP}_I}{\text{RFP}_I - \text{GFP}_I}$$

**Quantifizierung des  $E_{\text{ein}}$ -Proteinanteils:** Der  $E_{\text{ein}}$ -Proteinanteil ( $X_{\text{Ein}}$ ) wird über die GFP- und RFP-Fluoreszenzintensitäten ( $\text{GFP}_I$  und  $\text{RFP}_I$ ) bestimmt. In Abhängigkeit des Spleißreporters repräsentiert  $\text{GFP}_I$  die  $E_{\text{ein}}$ -Isoform, während  $\text{RFP}_I$  die  $E_{\text{ein}}$ - und  $E_{\text{aus}}$ -Isoform darstellt.

Im Zuge der Quantifizierung von Proteinisoformen mittels FACS wurden drei unabhängige biologische Replikate angefertigt. Für die Signifikanzprüfung unabhängiger Proben mit beliebig vielen Messgrößen kam ein zweiseitiger t-Test nach Student zur Anwendung.

#### 4.5.2 Fluoreszenzmikroskopie

Die Fluoreszenzmikroskopie wurde neben dem FACS zur Analyse des Spleißgeschehens auf Proteinebene eingesetzt. Eukaryotische Zellen wurden mit einer Zelldichte von  $3 \times 10^5$  Zellen pro Well einer Sechs-Well-Platte (BD Falcon, Heidelberg) auf sterile Deckgläschen ausgesät und transfiziert (Abschn. 4.3.1-3). Nach 48 h Transfektionsdauer wurden die Zellen dreimal mit kaltem PBS gewaschen und 30 min bei 4 °C mit Paraformaldehyd-PBS-Lösung (4 % (v/v) Paraformaldehyd (Merck KGaA, Darmstadt)) fixiert. Paraformaldehyd vernetzt zelluläre Proteine irreversibel, wodurch Fluoreszenzpräparate für längere Zeit konserviert werden. Ein Ausbleichen der Fluorophore wurde weitgehend unterbunden, da alle folgenden Schritte unter Schwachlichtbedingungen durchgeführt wurden. Die Fixierungslösung wurde verworfen und Paraformaldehydrückstände durch dreimaliges Waschen mit kalter PBS-Lösung entfernt. Die Fertigstellung der Fluoreszenzpräparate erfolgte, indem 20 µl Glycerol-PBS-Lösung (50 % (v/v) Glycerol) auf einen Objektträger pipettiert und ein mit Zellen bewachsenes Deckgläschen darauf luftblasenfrei fixiert wurde. Mit Lackringen konnten die Präparate fixiert und für maximal zwei Wochen konserviert werden.

Die Fluoreszenzaufnahmen wurden mit einem Laser-Scanning-Mikroskop (*Leica TCS SP5* von Leica Microsystems, Wetzlar) gemacht. Die Detektion der Fluorophore AcGFP1 und DsRed-Express erfolgte bei 488 nm Anregung mit dem Einsatz von *EGFP*- (495-539 nm) und *DsRed*- (578-620 nm) Filtern.

## 4.6 Biometrische Datenauswertung

Die Datenauswertung der DNA-, RNA- und Proteinanalysen erfolgte unter Verwendung der Programme *Microsoft Excel 2007* (Microsoft) und *R* (R Development) (Index et al. 2003).

Zur Charakterisierung identifizierter spleißregulatorischer Motive wurden Assoziationsstudien mit bekannten, mathematisch beschriebenen spleißregulatorischen Faktoren wie GC-Gehalt, RNA-Sekundärstruktur, RNA-Bindeproteine etc. durchgeführt. Die Bestimmung der Stärke der Assoziationen erfolgte durch Berechnung der Effektgröße (d) und Effektkorrelation (r) eines Zwei-Stichproben-t-Tests für unabhängige Stichproben mit folgenden Formeln:

$$d = \frac{t(n_E + n_S)}{(\sqrt{df})(\sqrt{(n_E + n_S)})}$$

**Berechnung der Effektgröße:** Die Effektgröße (d) berechnet sich aus dem t-Wert (t), den Freiheitsgraden (df) und dem Stichprobenumfang der Enhancer- ( $n_E$ ) und Silencer-Sequenzen ( $n_S$ ) des t-Tests.

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

**Berechnung der Effektkorrelation:** Die Effektkorrelation (r) berechnet sich aus dem t-Wert (t) und den Freiheitsgraden (df) des t-Tests. r hat einen minimalen Wert von 0 und einen maximalen Wert von 1.

Inwieweit die identifizierten spleißregulatorischen Faktoren den quantifizierten Einfluss auf die Spleißaktivität beschrieben, wurde mittels linearer Regression unter Verwendung des Programms *IBM SPSS Statistics 21* (IBM) ermittelt.

## 5 ERGEBNISSE

### 5.1 Das Spleißreportersystem

#### 5.1.1 Klonierung und Funktionalität des Reporters

Die Grundlage der Analyse von spleißregulatorischen Motiven (SRMs) bildete die Konstruktion des rHBBg-Reporters (Abb. 8). Dieses Plasmid besteht aus einem Fragment des  $\beta$ -Globingens (*HBB*), welches von den kodierenden Regionen der fluoreszierenden Proteine DsRed-Express in 5'-Richtung und AcGFP1 in 3'-Richtung im Expressionsvektor pAcGFP1-N1 flankiert wird. Das *HBB*-Genfragment (1418 bp) wurde als Analysegegenstand gewählt, da es, abgesehen von den nativen Start- und Stoppkodonen und den 5'- und 3'-UTRs, den vollständigen Genkontext (Exon 1-3) umfasst und eine klonierbare Größe besitzt.

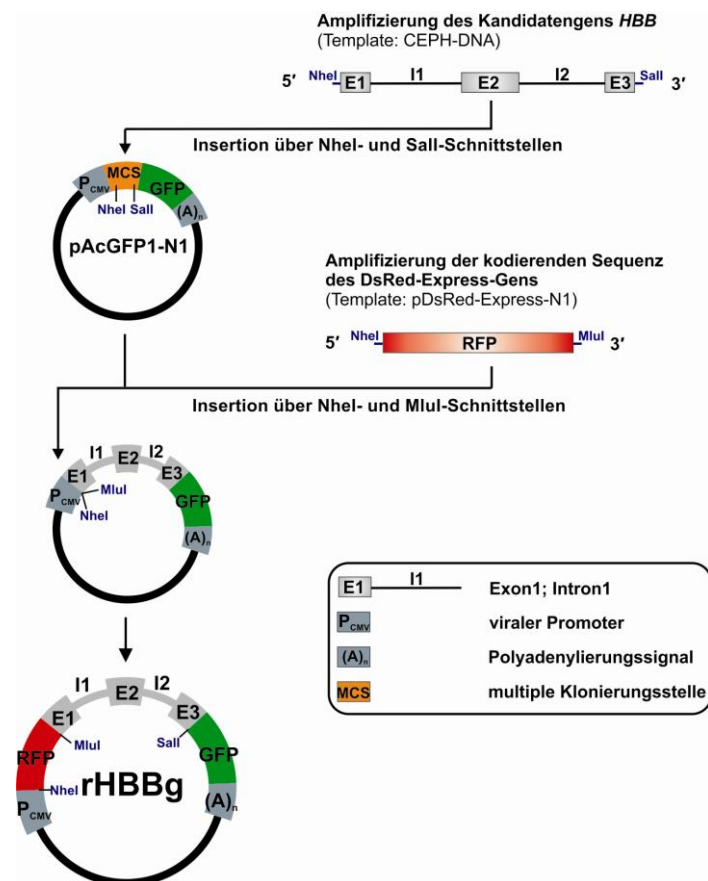
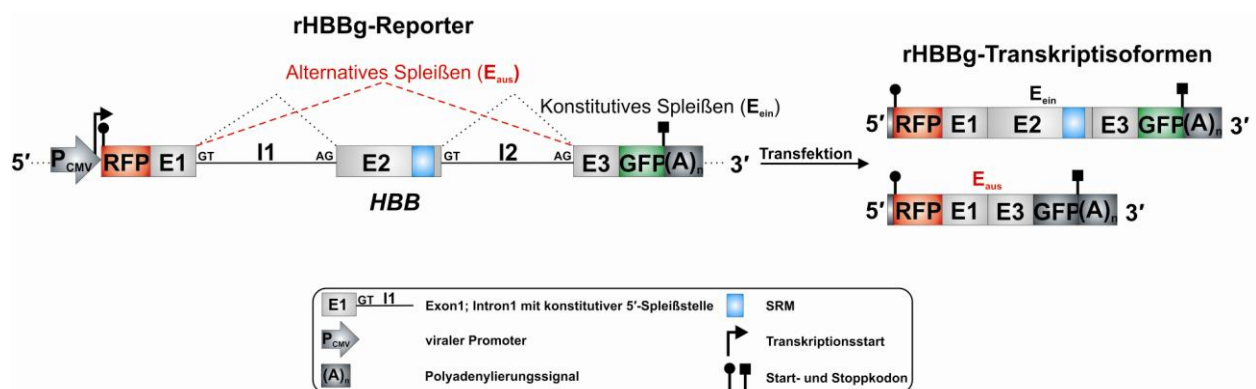


Abb. 8: Klonierungsschema des rHBBg-Reporters

Für zukünftige Analysen wurden neben dem Kandidatengen *HBB* auch *ACTB* (Exon 1-3), *APOC3* (Exon 2-4), *EIF3D* (Exon 7-9), *EPCAM* (Exon 2-4), *E2F4* (Exon 5-7), *LMNA* (Exon 10-12), *MAT1A* (Exon 7-9) und *PGM2L1* (Exon 10-12) in identischer Weise kloniert (Anhang I).

Die Konstruktion des rHBBg-Reporters erfolgte in Hinblick auf die Analyse von SRMs. So vermitteln SRMs, die das konstitutive prä-mRNA-Spleißen ( $E_{\text{ein}}$ ) des Reportertranskriptes fördern,

eine funktionelle Expression des rekombinanten Fusionsproteins RFP-HBB-GFP (rHBBg) in transfizierten eukaryotischen Zellen. Die Fluorophore GFP und RFP können mittels FACS bzw. Fluoreszenzmikroskopie gleichzeitig detektiert werden. Da das Exon 2 aus 223 bp besteht und ein Exon-übergreifendes Kodon am 5'-Ende enthält, wird im Fall eines alternativen Spleißereignisses ( $E_{aus}$ ) eine Leserahmenverschiebung bewirkt und bei ausbleibendem NMD ein verkürztes Protein synthetisiert. Demzufolge ist die GFP-Domäne des Fusionsproteins nicht mehr funktionell. Eine ausbleibende GFP- und RFP-Fluoreszenz ist ein Indiz für einen NMD-Effekt. Die schematische Darstellung der Funktionalität des rHBBg-Reporters befindet sich in Abbildung 9.



**Abb. 9: Schematische Darstellung der Funktionalität des Spleißreportersystems:** Spleißregulatorische Motive (SRMs) in der Nähe einer 5'-Spleißstelle (GT) vermitteln konstitutive ( $E_{ein}$ ) und alternative ( $E_{aus}$ ) Spleißreaktionen. Im Gegensatz zum konstitutiven Spleißen führt der Ausschluss von Exon 2 zu einer Leserahmenverschiebung und demzufolge zur Prozessierung eines Transkriptes mit einer nicht funktionellen GFP-Region in transfizierten eukaryotischen Zellen.

Nach der Transfektion des rHBBg-Reporters in eukaryotische Zellen können die  $E_{ein}$ - und  $E_{aus}$ -Isoformen auf Transkriptebeine mittels hochauflösender Fragmentlängenanalyse (CE-LIF) und auf Proteinebene mittels FACS-Analyse quantifiziert werden (Abschn. 4.4.5 und 4.5.1). Sind dabei die Mengenverhältnisse der Isoformen auf Transkript- und Proteinebene gleich, können zukünftig durch FACS-Analysen der Proteinisoformen schnell und kostengünstig Rückschlüsse auf die Transkriptisoformen gezogen werden (Schoft et al. 2007).

### 5.1.2 Funktionelle Reportervalidierung

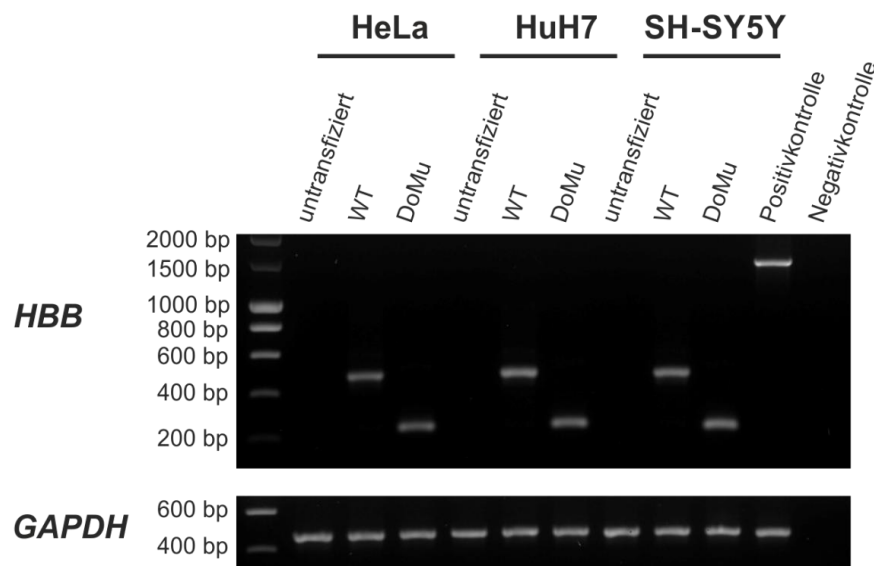
Zur funktionellen Validierung des Spleißreportersystems auf Transkript- und Proteinebene wurden die in den späteren Analysen verwendeten humanen Zelllinien HeLa, HuH7 und SH-SY5Y mit dem rHBBg-Reporter transfiziert.

- Funktionelle Reportervalidierung auf Transkriptebeine**

Zur Analyse des Spleißens auf Transkriptebeine wurde die Gesamt-RNA transfizierter Zellen isoliert, in cDNA umgeschrieben und RT-PCRs durchgeführt. Für die RT-PCRs kamen Primer



zur Anwendung, deren Produkt das zu analysierende Exon (Exon 2) und die flankierenden Exons (Exon 1 und Exon 3) überspannte. Um überprüfen zu können, ob das Reportersystem konstitutive und alternative Spleißereignisse prozessieren kann, wurde neben dem rHBBg-Wildtypplasmid, welches einen  $E_{\text{ein}}$  vermittelte, eine Mutante des rHBBg-Reporters mitgeführt. Die Mutante wies eine Punktmutation des Spleißdonors des *HBB*-Introns 2 auf (GT>CA), die mithilfe einer ortsgerichteten Mutagenese erzeugt wurde und die zum  $E_{\text{aus}}$  führte (Abb. 10).



**Abb. 10: Funktionelle Validierung des Spleißreportersystems auf Transkriptebe:** Das mit Ethidiumbromid gefärbte Agarosegel stellt die Ergebnisse der RT-PCRs dar, die zur Spleißanalyse von untransfizierten, rHBBg-Wildtypplasmid- (WT) und rHBBg-Donormutationsplasmid- (DoMu) transfizierten HeLa-, HuH7- und SH-SY5Y-Zellen angefertigt wurden. Während die  $E_{\text{ein}}$ -Isoform eine Länge von 438 bp aufweist, besitzt die  $E_{\text{aus}}$ -Isoform eine Länge von 215 bp. Der Einsatz einer einheitlichen Gesamt-RNA-Menge wurde durch den *GAPDH*-Nachweis erbracht. Im Fall von *HBB* diente rHBBg-Wildtypplasmid-DNA als Positivkontrolle. Die Dauer der Elongationsphase der RT-PCRs reichte aus, um das vollständige *HBB*-Gen (1418 bp) zu amplifizieren. Zur *GAPDH*-Analyse wurde HeLa-Zellen-cDNA als Positivkontrolle verwendet.

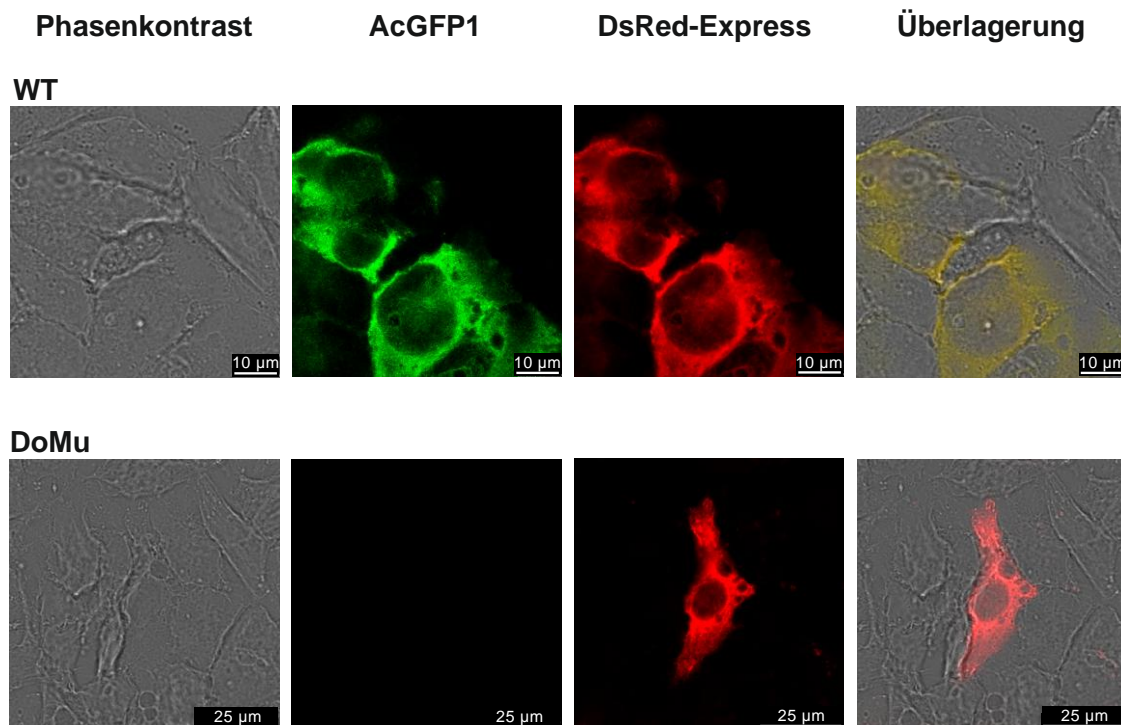
Die Bewertung der 5'-Spleißstellenstärke des *HBB*-Introns 2 wurde zuvor mit dem *MaxEntScan::score5ss for human 5' splice sites* Programm der MIT Christopher Burge Laboratory Homepage durchgeführt (Yeo & Burge 2004). Während bei einem maximalen Score von 10 und einem minimalen Score von -10 die native 5'-Spleißstelle einen Score von 9,25 erzielte, sank die Spleißstellenstärke bei dem alternativen CA-Donor auf -7,20 Punkte. Somit wurde ein Ausschluss des zentralen Exons von Beginn an wahrscheinlich und diese Punktmutation als Kontrolle eines alternativen Spleißereignisses ausgewählt.

Aus Abbildung 10 wird ersichtlich, dass *HBB* in den getesteten Zelllinien nicht endogen exprimiert wird und dementsprechend keine endogene Kontamination in den folgenden Analysen zu erwarten war.

#### • Funktionelle Reportervalidierung auf Proteinebene

Zur funktionellen Validierung des Spleißreportersystems auf Proteinebene wurde die Fluoreszenzmikroskopie genutzt. Während rHBBg-Wildtypplasmid-transfizierte eukaryotische Zellen ein

funktionelles rHBBg-Fusionsprotein synthetisierten, vermittelte die Donormutation des *HBB*-Introns 2 eine  $E_{\text{aus}}$ -Reaktion und demzufolge die Expression eines verkürzten Proteins mit lediglich einer funktionellen RFP-Domäne. In Abbildung 11 sind die Ergebnisse der Analyse in HeLa-Zellen dargestellt. Die Zelllinien HuH7 und SH-SY5Y wiesen die gleichen Fluoreszenzaktivitäten auf.



**Abb. 11: Funktionelle Validierung des Spleißreportersystems auf Proteinebene in HeLa-Zellen:** Mikroskopische Untersuchungen von HeLa-Zellen nach Transfektion mit dem rHBBg-Wildtypplasmid (WT) und rHBBg-Donormutationsplasmid (DoMu). **Phasenkontrast:** Darstellung der Zellen im Phasenkontrast. **AcGFP1:** Darstellung der Zellen unter Verwendung des *EGFP*-Filters. **DsRed-Express:** Darstellung der Zellen unter Verwendung des *DsRed*-Filters. **Überlagerung:** Überlagerung der drei Aufnahmen. Die Ergebnisse zeigen, dass das WT eine funktionelle rHBBg-Fusionsproteinexpression vermittelt und das DoMu durch Auslösen einer  $E_{\text{aus}}$ -Reaktion lediglich eine DsRed-Express-Fluoreszenzaktivität erzeugt.

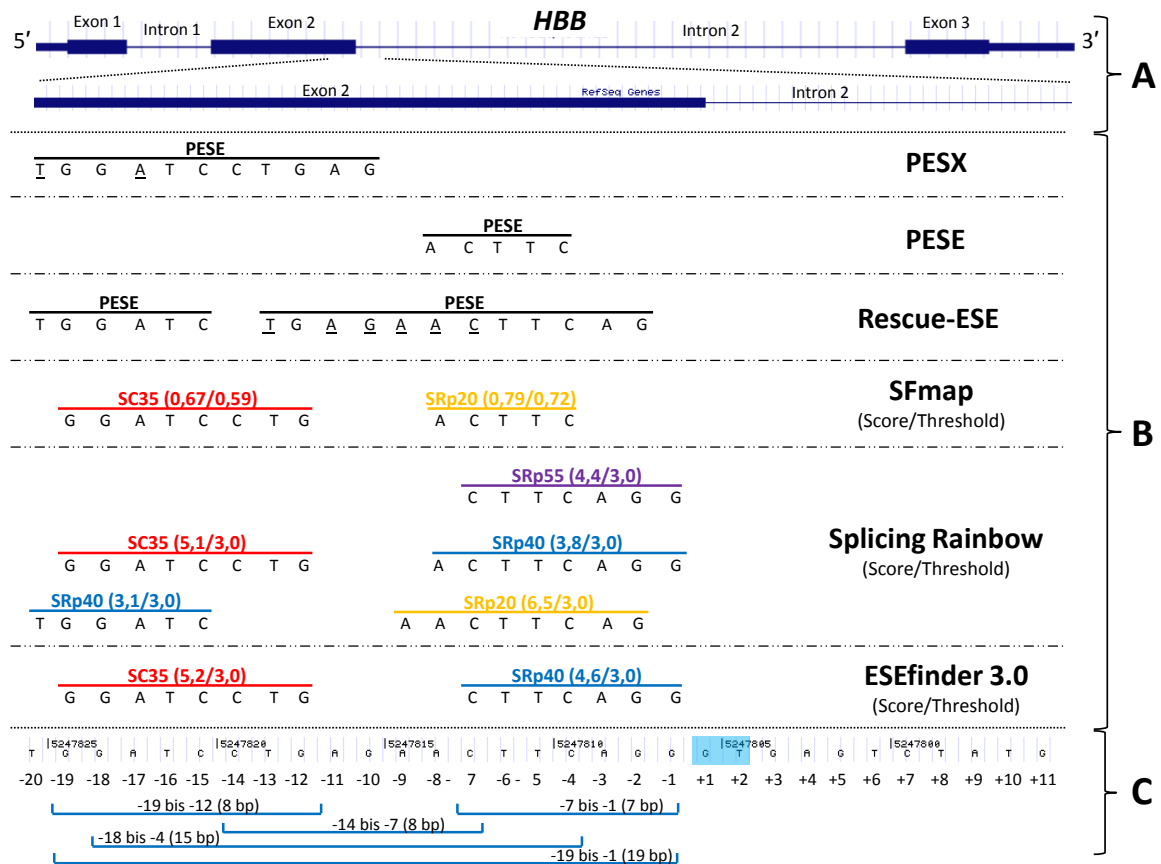
Der Vergleich der Transkriptgrößen und Fluoreszenzen, die vom rHBBg-Wildtypplasmid und rHBBg-Donormutationsplasmid erzeugt wurden, zeigt, dass das Spleißreportersystem konstitutive und alternative Spleißreaktionen darstellen kann und somit zur Analyse von SRMs geeignet ist.

### 5.1.3 Identifizierte spleißregulatorische Loci

Die Identifizierung von SRMs setzt die Kenntnis über deren Locus im *HBB*-Gen voraus. Daher wurden spleißregulatorische Loci mit einer maximalen Entfernung von 20 nt zur 5'-Spleißstelle des *HBB*-Introns 2 bioinformatisch identifiziert und anschließend experimentell auf Funktionalität getestet. Da intronische spleißregulatorische Loci lediglich mittels ineffizienter Lariat-PCRs analysiert werden können, wurde in dieser Arbeit der Schwerpunkt auf die Analyse von exonischen Loci gelegt (Gao et al. 2008).

### • Bioinformatisch identifizierte spleißregulatorische Loci

Die computerbasierte Detektion exonischer SRMs am 3'-Ende des zentralen *HBB*-Exons wurde unter Verwendung von sechs verschiedenen Online-Vorhersage-Programmen durchgeführt, deren Datengrundlage bereits im Abschnitt 3.6 beschrieben wurde. Während die Programme *ESEfinder 3.0*, *Splicing Rainbow* und *SFmap* Spleißfaktor-spezifische RNA-Bindestellen angeben, werden von den Programmen *RESCUE-ESE*, *PESE* und *PESX* putative exonische SRMs vorhergesagt (Fairbrother et al. 2002; Cartegni et al. 2003; Zhang & Chasin 2004; Stamm et al. 2006; Goren et al. 2006; Paz et al. 2010). Die bioinformatisch identifizierten und als relevant eingestuften spleißregulatorischen Loci sind in Abbildung 12 dargestellt.



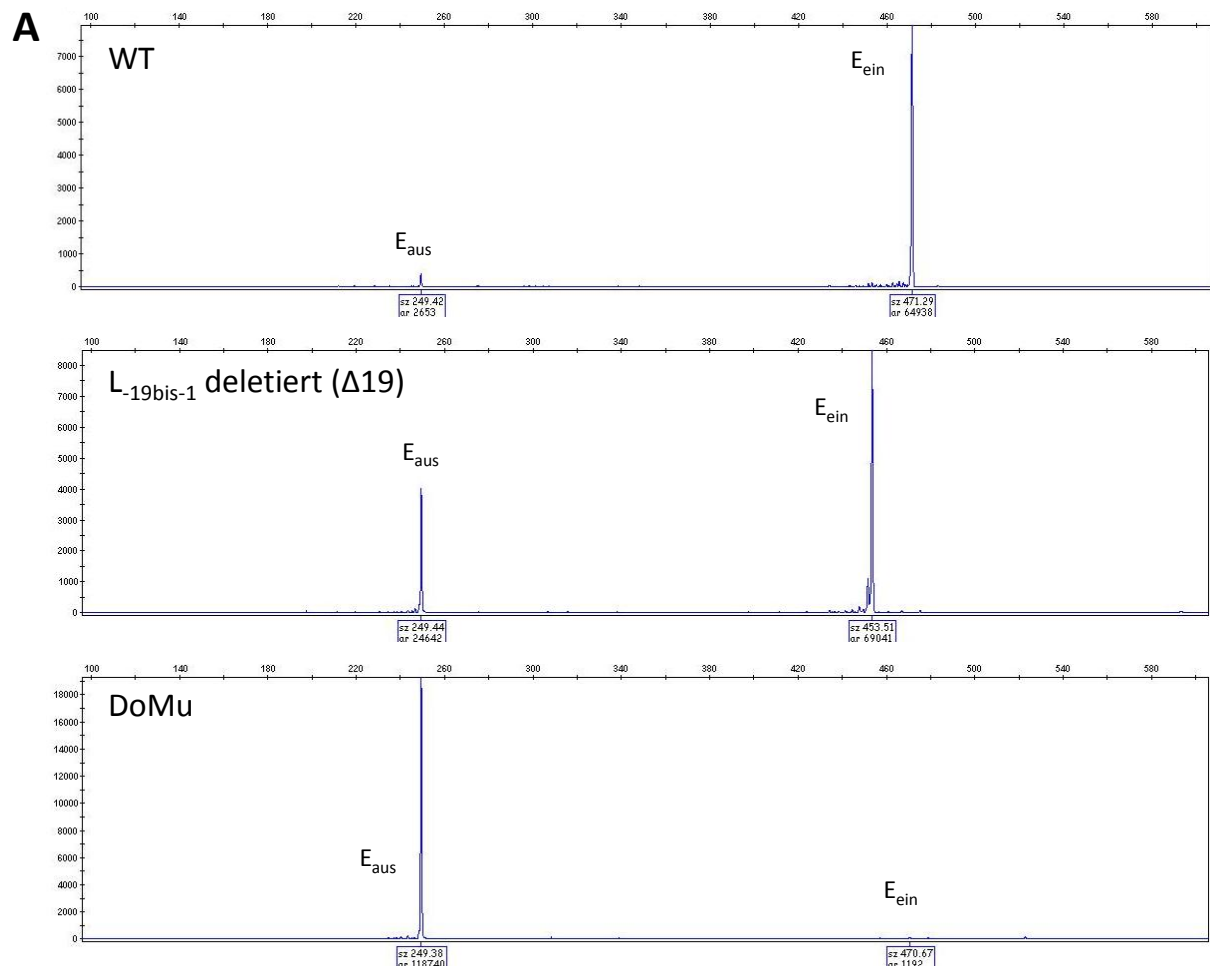
**Abb. 12: Bioinformatisch identifizierte spleißregulatorische Loci:** **A:** Schematische Darstellung des *HBB*-Gens im UCSC-Genome-Browser. **B:** Die bioinformatisch identifizierten spleißregulatorischen Loci der einzelnen Vorhersageprogramme sind korrespondierend zum Gen angeordnet. Die Bindepartner der identifizierten Motive werden von den Programmen *ESEfinder 3.0*, *Splicing Rainbow* und *SFmap* angegeben. Die Programme *Rescue-ESE*, *PESE* und *PESX* identifizieren putative exonische spleißregulatorische Motive (PESE). **C:** Die Loci mit den Positionen -19 bis -12, -14 bis -7, -7 bis -1, -18 bis -4 und -19 bis -1 wurden aufgrund der Ergebnisse aus B als putative spleißregulatorische Loci identifiziert. Die hellblau hinterlegten Basen kennzeichnen die 5'-Spleißstelle des *HBB*-Introns 2.

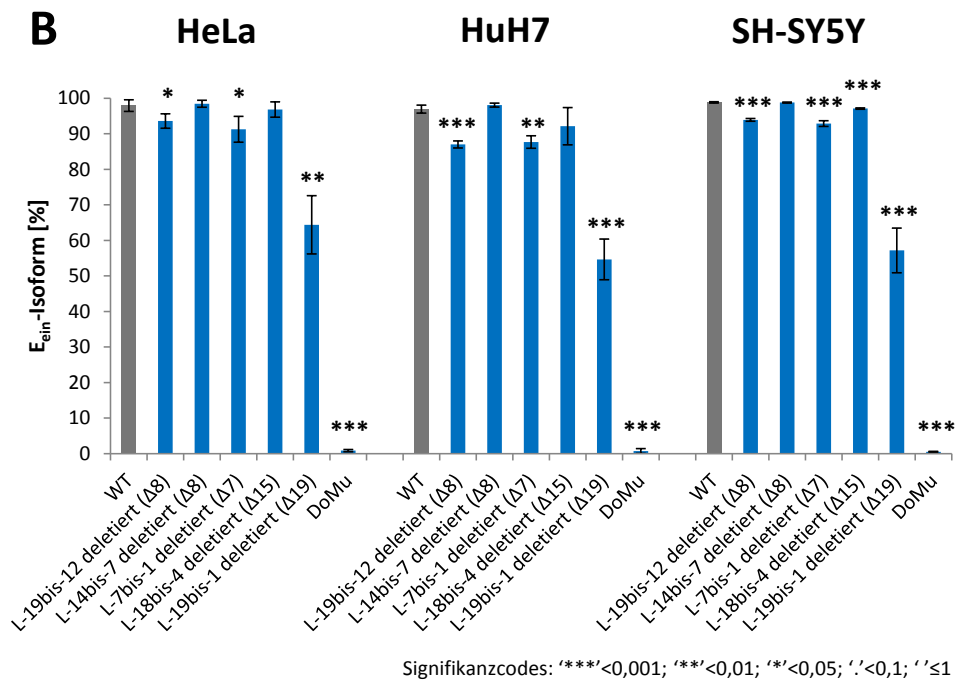
Aus dem Ergebnis der bioinformatischen Vorhersage spleißregulatorischer Loci in 5'-Spleißstellennähe des *HBB*-Introns 2, wurden die in Abbildung 12C dargestellten Loci mit den Positionen -19 bis -12 ( $L_{-19\text{bis}-12}$ ), -14 bis -7 ( $L_{-14\text{bis}-7}$ ), -7 bis -1 ( $L_{-7\text{bis}-1}$ ), -18 bis -4 ( $L_{-18\text{bis}-4}$ ) und -19 bis -1 ( $L_{-19\text{bis}-1}$ ) als Konsens der sechs verschiedenen Vorhersageprogramme abgeleitet und mussten nun experimentell auf Funktionalität überprüft werden.

- **Experimentell identifizierte spleißregulatorische Loci**

Mithilfe von Deletionsexperimenten wurden die fünf ausgewählten *HBB*-Loci auf spleißregulatorische Funktionalität untersucht. Zur Deletion der Loci wurden ortsgerichtete Mutagenesen durchgeführt. Dementsprechend wiesen die mutierten rHBBg-Plasmide ein verkürztes Exon 2 auf. Zur Beibehaltung des GFP-Leserahmens wurden ebenfalls mit ortsgerichteten Mutagenesen bis zu 2 bp der verbliebenen multiplen Klonierungsstelle stromabwärts des *HBB*-Exons 3 deletiert. Der neu entstandene Spleißkontext erreichte einen Basenunterschied zum rHBBg-Wildtypplasmid (Wildtypabweichung, WA) von mindestens 50 %. Erfolgte auf diese Weise die Inaktivierung eines hochregulativen Motivs, so wurde im Vergleich zum Wildtyp (WT) ein signifikanter Unterschied im  $E_{\text{ein}}/E_{\text{aus}}$ -Isoformenverhältnis auf Transkript- und Proteinebene detektiert. Zur Quantifizierung der  $E_{\text{ein}}$ - und  $E_{\text{aus}}$ -Spleißvarianten wurden die Deletionskonstrukte in HeLa-, HuH7- und SH-SY5Y-Zellen transfiziert. Als Kontrolle einer effektiven  $E_{\text{aus}}$ -Reaktion wurde das rHBBg-Donormutationsplasmid mitgeführt.

Zunächst wurden die  $E_{\text{ein}}/E_{\text{aus}}$ -Isoformenverhältnisse durch Messung der Transkriptlängen bestimmt. Hierfür wurde die Gesamt-RNA transfizierter Zellen isoliert, in cDNA umgeschrieben und RT-PCRs durchgeführt. Für die RT-PCRs kamen Primer zur Anwendung, deren Produkt das zu analysierende Exon (Exon 2) und die flankierenden Exons (Exon 1 und Exon 3) überspannte. Im Anschluss erfolgten CE-LIF-Analysen (Abb. 13 und Anhang III).



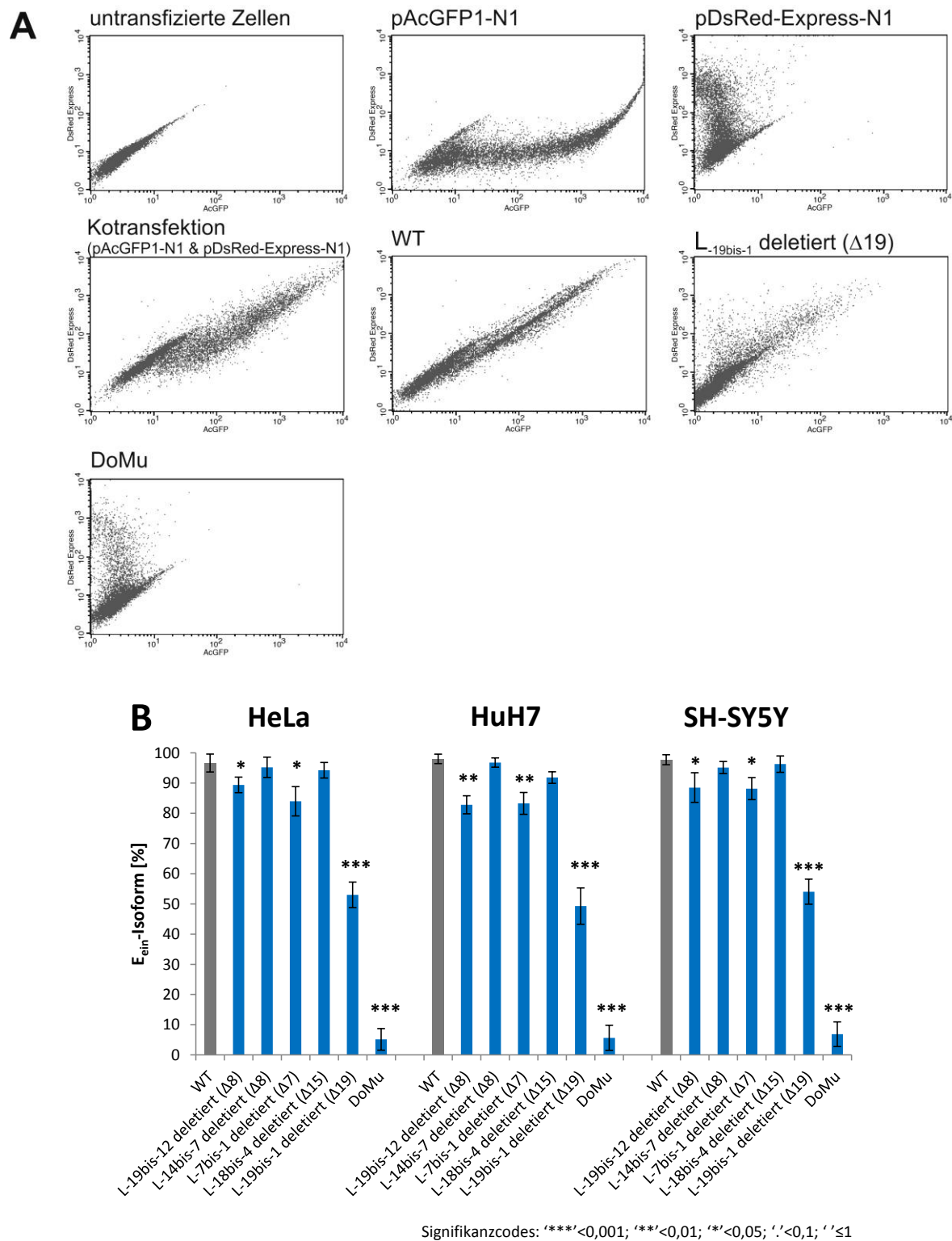


**Abb. 13: CE-LIF-basierte Bewertung der bioinformatisch identifizierten spleißregulatorischen *HBB*-Loci:** **A:** Elektropherogramme der rHBBg-Wildtypplasmid- (WT), rHBBg-Deletionsplasmid- (L-19bis-1 deletiert) und rHBBg-Donormutationsplasmid- (DoMu) transfizierten HeLa-Zellen zur Veranschaulichung einer CE-LIF-Analyse. Entlang der Abszisse sind die Transkriptlängen [nt] aufgetragen. Die Ordinate stellt die Fluoreszenzintensitäten dar. Die Peak-Position (sz) und -Fläche (ar) charakterisieren die E<sub>aus</sub>- und E<sub>ein</sub>-Transkriptisoformen. **B:** Die Ordinatenachse repräsentiert die prozentualen Verteilungen der E<sub>ein</sub>-Isoformen, einschließlich Standardabweichungen. Signifikante Änderungen der E<sub>ein</sub>-Isoformen der Mutationsplasmide, im Vergleich zum WT, wurden mittels zweiseitigem t-Test nach Student berechnet. „Δ“ symbolisiert die Anzahl deletierter Basenpaare des *HBB*-Exons 2.

Die CE-LIF-Ergebnisse zeigen, dass die Deletion putativer spleißregulatorischer *HBB*-Loci sowie die Mutation des Spleißdonors des *HBB*-Introns 2 im Vergleich zum WT zu einer Zunahme der E<sub>aus</sub>-Transkriptisoform führten und demzufolge eine alternative Spleißreaktion verstärkten. Die Positivkontrolle, die Mutation des Spleißdonors, besaß dabei in den getesteten Zelllinien den größten spleißregulatorischen Effekt und führte in durchschnittlich 99,3 % ± 0,2 % der Fälle zu einem E<sub>aus</sub>.

In allen getesteten Zelllinien wiesen die Positionen L-19bis-12, L-7bis-1 und L-19bis-1 einen signifikanten Einfluss auf die Spleißaktivität der 5'-Spleißstelle des *HBB*-Introns 2 auf, wobei der Effekt in HuH7-Zellen am größten war. Hier zeigt der Vergleich mit dem WT, dass durch die Deletion der Bereiche L-19bis-12 bis zu 10,0 % ± 0,1 %, L-7bis-1 bis zu 9,3 % ± 2,4 % und L-19bis-1 bis zu 42,3 % ± 6,6 % mehr E<sub>aus</sub>-Transkriptisoform erzeugt wurde. Den Regionen L-14bis-7 und L-18bis-4 konnte keine zelllinienübergreifende spleißregulatorische Funktion nachgewiesen werden. Jedoch wird aus dem Vergleich von L-19bis-1 und L-18bis-4 deutlich, dass die drei benachbarten Basen stromaufwärts des Spleißdonors verstärkt in das Spleißgeschehen eingriffen.

Um die quantifizierten Spleißvarianten auf Proteinebene zu bestätigen und die funktionelle Validierung des Spleißreportersystems abzuschließen, erfolgte im Anschluss die Untersuchung von E<sub>ein</sub>- und E<sub>aus</sub>-Spleißvarianten transfizierter Zellen mittels FACS-Analysen (Abb. 14 und Anhang IV).



**Abb. 14: FACS-basierte Bewertung der bioinformatisch identifizierten spleißregulatorischen *HBB*-Loci:** **A:** FACS-Analyse zur Gegenüberstellung der Fluoreszenzintensitäten von AcGFP1 (Abszisse) und DsRed-Express (Ordinate) am Beispiel untransfizierter, pAcGFP1-N1-, pDsRed-Express-N1-, ko-, rHBBg-Wildtypplasmid- (WT), rHBBg-Deletionsplasmid- (L<sub>19bis-1</sub> deletiert) und rHBBg-Donormutationsplasmid- (DoMu) transfizierter HeLa-Zellen. **B:** Die Ordinate repräsentiert die prozentualen Verteilungen der  $E_{\text{ein}}$ -Isoformen, einschließlich Standardabweichungen. Signifikante Änderungen der  $E_{\text{ein}}$ -Isoformen der Mutationsplasmide, im Vergleich zum WT, wurden mittels zweiseitigem t-Test nach Student berechnet. „ $\Delta$ “ symbolisiert die Anzahl deletierter Basenpaare des *HBB*-Exons 2.

Wie auch die CE-LIF-Analyse zuvor, zeigt das FACS-Experiment, dass die Deletionen von L<sub>19bis-12</sub>, L<sub>7bis-1</sub> und L<sub>19bis-1</sub> und die Mutation des Spleißdonors in den drei getesteten Zelllinien



einen signifikanten spleißregulatorischen Einfluss auf die 5'-Spleißstellenerkennung des *HBB*-Introns 2 besaßen. Der Vergleich der CE-LIF- und FACS-Ergebnisse verdeutlicht zudem, dass beide Methoden gleiche  $E_{\text{ein}}/E_{\text{aus}}$ -Isoformenverhältnisse quantifizierten ( $p > 0,05$ ; zweiseitiger t-Test nach Student) (Anhang V). Demzufolge konnten mit zwei unabhängigen Methoden die Bereiche  $L_{-19\text{bis}-12}$ ,  $L_{-7\text{bis}-1}$  und  $L_{-19\text{bis}-1}$  als spleißregulatorische Loci identifiziert, gleiche Mengenverhältnisse der Isoformen auf Transkript- und Proteinebene ermittelt und letztendlich die Spleißreportervalidierung abgeschlossen werden.

#### • Wahl der zu analysierenden Loci

Zur Identifizierung von SRMs in 5'-Spleißstellennähe wurden die spleißregulatorischen Bereiche  $L_{-19\text{bis}-12}$ ,  $L_{-7\text{bis}-1}$  und  $L_{-19\text{bis}-1}$  des *HBB*-Introns 2 ausgewählt. Diese sollten nun auf der Grundlage des rHBBg-Reporters randomisiert mutagenisiert werden. Da die randomisierte Mutagenese des Bereichs  $L_{-19\text{bis}-1}$  mit  $4^{19}$  möglichen Motiven eine zu hohe Komplexität aufweist, wurden in dieser Arbeit die Positionen  $L_{-19\text{bis}-12}$  und  $L_{-7\text{bis}-1}$  analysiert.

Aufgrund dessen, dass *trans*-Faktoren nicht nur mutierte, sondern ebenso flankierende Basen binden, ergab sich eine hohe Motivkomplexität. Beispielsweise vermittelte die Mutation von 8 bp 15 mögliche SRM-Positionen. So führte die Mutagenese der Bereiche  $L_{-19\text{bis}-12}$  und  $L_{-7\text{bis}-1}$  zur Analyse der Positionen -26 bis -5 und -13 bis +6 des *HBB*-Introns 2 (Abb. 15).

	-26	-25	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6
WT	T	G	C	A	C	G	T	G	G	A	T	C	C	T	G	A	G	A	A	C	T	T	C	A	G	G	g	t	g	a	g	t
1	T	G	C	A	C	G	T	N												T	G	A	G	A	A	N						
2																				G	A	G	A	A	N	N						
3																				C	A	C	G	T	N	N	N					
4																				A	C	G	T	N	N	N	N					
5																				C	G	T	N	N	N	N	N					
6																				G	T	N	N	N	N	N	N					
7																				T	N	N	N	N	N	N	N					
8																				N	N	N	N	N	N	N	N	g				
9																				N	N	N	N	N	N	N	N	g	t			
10																				N	N	N	N	N	N	N	N	g	t	g		
11																				N	N	N	N	N	N	N	N	g	t	g	a	
12																				N	N	N	N	N	N	N	N	g	t	g	a	g
13																				N	N	N	N	N	N	N	N	g	t	g	a	g
14																				N	N	N	N	N	N	N	N	g	t	g	a	g
15																				N	N	N	N	N	N	N	N	g	t	g	a	g

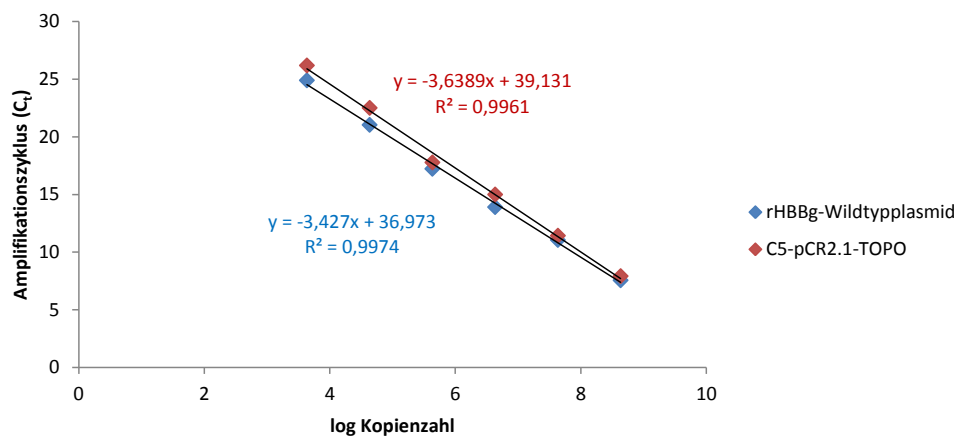
**Abb. 15: Motivkomplexität:** Die randomisierten Mutagenesen der Positionen  $L_{-19\text{bis}-12}$  und  $L_{-7\text{bis}-1}$  des nativen *HBB*-Introns 2 (WT) erzeugen dynamische 22-mer ( $L_{-26\text{bis}-5}$ ) und 19-mer ( $L_{-13\text{bis}+6}$ ) Motivkontexte. Groß- und Kleinbuchstaben unterscheiden Basen exonischen und intronischen Ursprungs. Die hellblau hinterlegten Basen kennzeichnen die 5'-Spleißstelle des *HBB*-Introns 2.

Da die zu mutierenden Sequenzabschnitte  $L_{-19\text{bis}-12}$  distal und  $L_{-7\text{bis}-1}$  proximal zur 5'-Spleißstelle des *HBB*-Introns 2 lokalisiert sind, werden diese der Einfachheit halber in den folgenden Abschnitten als distale (DL) und proximale Loci (PL) bezeichnet, die einen Spleißkontext in den Positionen -26 bis -5 (DL) und -13 bis +6 (PL) beschreiben.

### 5.1.4 Bestimmung der Transfektionseffizienz

Infolgedessen, dass das Spleißreportersystem eine Analyse von  $E_{\text{ein}}$ - und  $E_{\text{aus}}$ -Reaktionen auf Proteinebene gewährleistet, könnte vor der kostenintensiven Tiefensequenzierung der Transkripte eine FACS-vermittelte Vorselektion konstitutiv gespleißter Transkripte stattfinden und somit eine Anreicherung funktioneller SRMs erzielt werden. Voraussetzung für eine Vorselektion wäre, dass sich in jeder transient transfizierten Zelle nur ein rHBBg-Plasmid befindet.

Um dies zu klären, wurde die Plasmidkopienzahl pro transfizierter Zelle mittels qRT-PCR bestimmt. Die Berechnung der Kopienzahl erfolgte dabei durch das Erstellen einer Plasmidstandardkurve aus definierten Plasmidkopienzahlen und deren zugehörigen  $C_t$ -Werten. Zum Nachweis der Existenz eines diploiden Chromosomensatzes wurden 149 bp des C5-Gens in den pCR2.1-TOPO-Vektor kloniert und ebenfalls analysiert (Abb. 16 und Anhang VI).



**Abb. 16: Plasmidstandardkurve zur Berechnung der Plasmidkopienzahl:** Entlang der Abszisse sind die logarithmierten Plasmidkopienzahlen dargestellt und die Ordinate repräsentiert die zugehörigen Amplifikationszyklen.

Im Anschluss wurde die Gesamt-DNA aus rHBBg-Wildtypplasmid-transfizierten HeLa-Zellen isoliert und ein 155 bp großes RFP-Fragment sowie ein 149 bp großes C5-Genfragment amplifiziert. Bei einem Einsatz von 5 ng Gesamt-DNA als Template erfolgte bei einem durchschnittlichen Gewicht von 0,006 ng DNA pro HeLa-Zelle die Analyse von rund 833 Zellen (Joshi et al. 2008). Mithilfe der Plasmidstandardkurven wurden die Kopienzahlen berechnet (Tab. 2).

**Tab. 2: Berechnete Plasmid- und Genkopienzahlen pro diploider HeLa-Zelle**

Plasmid/Gen	$C_t$ -Wert	log Kopienzahl	Kopienzahl	Kopienzahl pro Zelle	durchschnittliche Kopienzahl pro Zelle
rHBBg-Wildtypplasmid	16,14	6,0791	1199714,5079	1439,6574	1240,64 ± 213,50
	16,66	5,9273	845944,1633	1015,1330	
	16,33	6,0236	1055931,7124	1267,1181	
Negativkontrolle (H <sub>2</sub> O)	>36	-	-	-	-
C5	27,02	3,3282	2129,1369	2,5550	2,20 ± 0,31
	27,42	3,2183	1653,0282	1,9836	
	27,37	3,2320	1706,1636	2,0474	
Negativkontrolle (H <sub>2</sub> O)	>36	-	-	-	-



Tabelle 2 kann entnommen werden, dass sich im Durchschnitt 1240 rHBBg-Wildtypplasmide in einer transient transfizierten HeLa-Zelle befanden. Somit konnte eine FACS-Vorselektion zur Analyse von SRMs nicht stattfinden. Die Gültigkeit der Ergebnisse wurde durch den Nachweis der Existenz eines diploiden Chromosomensatzes, mit einer C5-Genkopienzahl von  $2,20 \pm 0,31$ , erbracht.

## 5.2 Etablierung einer randomisierten Mutagenese

Für eine Identifizierung von SRMs war Voraussetzung, eine Methode zu etablieren, mit deren Hilfe die spleißregulatorischen Bereiche DL und PL des rHBBg-Reporters randomisiert mutagenisiert werden konnten.

### 5.2.1 Herstellung randomisierter Mutagenesebanken

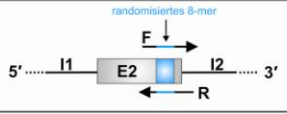
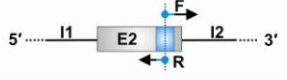
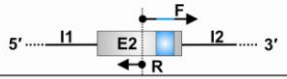
Im Rahmen dieser Doktorarbeit erfolgte die Etablierung einer randomisierten Mutagenese mithilfe der PCR-Technik. Insgesamt wurden drei verschiedene PCR-Ansätze zur Herstellung von randomisierten Mutagenesebanken am Beispiel des DL-Reporterbereichs durchgeführt. Eine zufallsverteilte Mutation von acht Basenpaaren erzielte dabei  $4^8$  Motivmöglichkeiten. Demzufolge bestand eine DL-Mutagenesebank aus 65536 verschiedenen Reportermolekülen, die sich in einem 8 bp langen Motiv unterschieden. Die Diversität der Mutagenesebanken wurde im Anschluss mittels Sanger-Sequenzierung ermittelt.

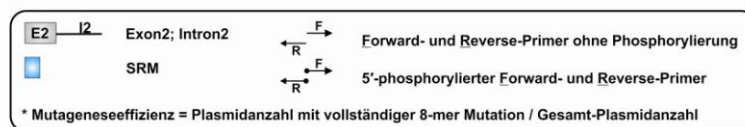
Der Ansatz A beruhte auf ortsgerichteten Mutagenesen mit zueinander komplementären Forward- und Reverse-Primern, die das gesamte, zufällig zusammengesetzte 8-mer Motiv beinhalten. Nach der Transformation der mutierten Plasmide in *E.coli* wurden einzelne Klone ausgewählt und sequenziert.

Die Ansätze B und C basierten auf einer PCR mit 5'-phosphorylierten, nicht zueinander komplementären Primer-Paaren. Der Unterschied zwischen beiden Ansätzen bestand in der Primer-Architektur. Während die Forward- und Reverse-Primer des Ansatzes B jeweils 4 nt des randomisierten 8-mer Motivs besaßen, war im Forward-Primer des Ansatzes C die vollständige 8-mer Mutationskassette lokalisiert. Die PCR-Produkte wurden anschließend mittels Agarose-Gelextraktion aufgereinigt. Die nächsten Arbeitsschritte beinhalteten eine DpnI-Restriktion, eine weitere Gelextraktion sowie eine Zirkularisierung der 5'-phosphorylierten PCR-Produkte. Zur Analyse der Diversität wurden die hergestellten Mutagenesebanken in *E.coli*-Zellen transformiert und bis zu 184 Klone sequenziert.

In Tabelle 3 sind die berechneten Mutageneseeffizienzen der PCR-Ansätze A, B und C aufgelistet.

Tab. 3: Mutageneseeffizienzen der PCR-Ansätze zur Herstellung einer Mutagenesebank

PCR-Ansatz	sequenzierte Plasmidanzahl				Mutagenese-effizienz* [%]
	vollständige 8-mer Mutation	<i>HBB</i> -Wildtypmotiv	unvollständige 8-mer Mutation	gesamt	
<b>A</b> 	123	2	51	176	69,99
<b>B</b> 	131	1	41	173	75,72
<b>C</b> 	177	1	6	184	96,20



Mit einer Mutageneseeffizienz von 96,20 % erwies sich das PCR-Modell C als am effektivsten und wurde in dieser Arbeit zur Herstellung einer randomisierten Mutagenesebank genutzt. Die geringe Anzahl des WT's war ein Indiz für eine nahezu vollständige Mutagenese.

### 5.2.2 Anreicherung der randomisierten Mutagenesebank

Um ausreichend Plasmid für nachfolgende Analysen zur Verfügung zu haben, musste eine Methode zur Anreicherung von Mutagenesebanken etabliert werden. Insgesamt wurden zwei Methoden für die Anreicherung etabliert und miteinander verglichen.

Zum einen wurde die benötigte Plasmidmenge durch die parallele Anfertigung von 65 PCR-basierten Mutagenesen erreicht, die im Laufe des Herstellungsprozesses zunehmend vereinigt wurden. Zum anderen wurde die Vervielfältigung einer einzigen Mutagenese durch Transformation des PCR-Produktes in *E.coli*-Zellen getestet. Da für die Anreicherung in *E.coli* nur wenige PCR-Schritte erforderlich waren, besaß dieses Anreicherungsverfahren einen technischen und finanziellen Vorteil. Dennoch musste überprüft werden, ob die beiden Methoden die Komplexität der Mutagenesebanken beeinflussen.

Für den Vergleich der Methoden wurde der Reporterbereich L<sub>19bis-1</sub> (19-mer) des *HBB*-Introns 2 randomisiert mutagenisiert. Da die randomisierte Mutation von 19 bp theoretisch ca.  $2,75 \times 10^{11}$  Motivmöglichkeiten erzielt und in diesem Experiment eine Tiefensequenzierung von 40 Mio. Plasmiden stattfand, konnte keine 100 %ige Motivabdeckung gewährleistet werden. Dementsprechend hätte im Fall einer ausreichenden Komplexität der Mutagenesebanken jedes Motiv nur einmal detektiert werden dürfen.

Die mittels PCR-Methode angereicherte Mutagenesebank wurde direkt zur Tiefensequenzierung eingesetzt. Im Gegensatz dazu musste bei der *E.coli*-Methode vor der Sequenzierung die Mutagenesebank aus den transformierten Bakterien isoliert werden.

Die sequenzierten Fragmente (engl. *reads*) wurden mit definierten Referenzsequenzen abgeglichen (aligniert). Dabei entsprach jeder Read einem sequenzierten Plasmid und demzufolge einem sequenzierten 19-mer Motiv. Auf diese Weise konnte für jedes erwartete 19-mer Motiv eine spezifische Read-Anzahl bestimmt werden (Tab. 4).

**Tab. 4: Vergleich der Methoden zur Anreicherung von Mutagenesebanken**

<b>Tiefensequenzierung</b>	<b>PCR-Methode</b>	<b><i>E.coli</i>-Methode</b>
Gesamt-Read-Anzahl	39.034.622	40.097.903
• Anzahl nicht alignierter Reads	2.331.090	2.601.901
• Anzahl alignierter Reads mit vollständiger 19-mer Mutation	36.703.532	37.496.002
➤ Read-Anzahl einmalig sequenzierter Motive	35.374.117	899.510
➤ Read-Anzahl des Wildtypmotivs	1.156.424	393.354

Anhand von Tabelle 4 wird ersichtlich, dass die mittels PCR synthetisierte Mutagenesebank 40x komplexer war als die in *E.coli* hergestellte DNA-Bank. Da von 36.703.532 sequenzierten rHBBg-Plasmiden 35.374.117 Plasmide ein unterschiedliches 19-mer Motiv beinhalteten, konnte davon ausgegangen werden, dass die Anreicherungs-methode durch PCR eine nahezu vollständige Motivkomplexität gewährleistete. Aus diesem Grund wurde diese Methode zur Anreicherung der randomisierten Mutagenesebanken verwendet. Der relativ hohe Anteil an Wildtypmotiv war ein Indiz für eine hohe Kontamination der PCR-Produkte mit Template-DNA. Da eine Kontamination mit rHBBg-Wildtypplasmid die Quantifizierung von SRMs verfälscht, erfolgte der Einsatz der rHBBg-Plasmide, deren DL- und PL-Reporterbereiche deletiert wurden (Abschn. 5.1.3), als Template der randomisierten Mutagenesen. Auf diese Weise konnte eine exakte Unterscheidung zwischen Template-DNA und Mutageneseprodukten erreicht werden.

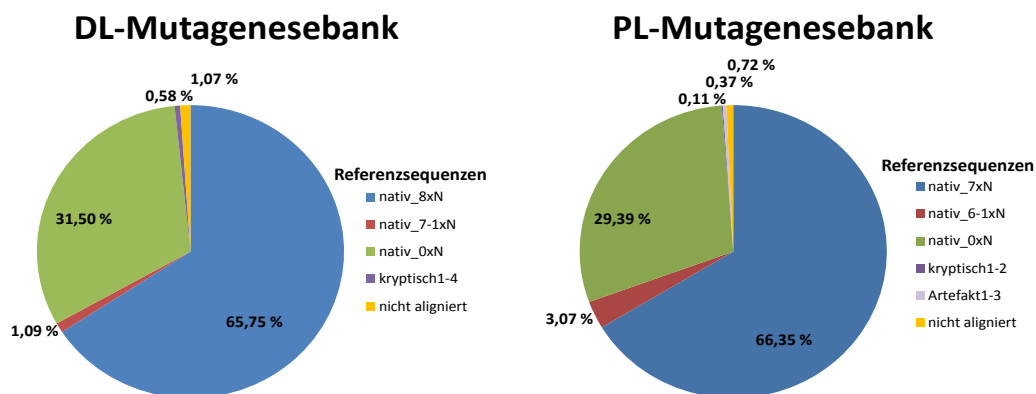
### 5.3 Identifizierung von spleißregulatorischen Motiven (SRMs)

Zur Identifizierung von SRMs wurden die identifizierten spleißregulatorischen DL- und PL-Reporterbereiche (Abschn. 5.1.3) mithilfe der etablierten Mutagenesemethode (Abschn. 5.2) randomisiert mutagenisiert. Die resultierenden Mutagenesebanken wurden anschließend in humane Zellen transient transfiziert, die RNA isoliert sowie in cDNA umgeschrieben und die Motiv-enhaltenden Regionen des Reportertranskriptes mittels RT-PCR amplifiziert. Eine Tiefensequenzierung der Mutagenesebanken und der spezifischen Transkript-Pools (RT-PCR-Produkte) ermöglichte eine Quantifizierung der transfizierten Plasmide und der daraus entstandenen Transkripte. Jedes sequenzierte Fragment (engl. *read*) entsprach dabei einem sequenzierten Plasmid bzw. Transkript. So wurde den erwarteten Motiven eine spezifische Read-Anzahl zugeordnet. Durch die Bestimmung des Verhältnisses der Read-Anzahlen von Transkript und korrespondierendem Plasmid konnte der Einfluss auf die Spleißaktivität jedes

einzelnen Motivs berechnet werden. Die Ergebnisse der jeweiligen Arbeitsschritte sind in den folgenden Abschnitten aufgeführt.

### 5.3.1 Sequenzabgleich der Mutagenesebanken

Die hergestellten DL- und PL-Mutagenesebanken wurden vor der Transfektion in eukaryotische Zelllinien mittels Tiefensequenzierung überprüft. Zur Amplifikation der zu untersuchenden Abschnitte wurden Forward- und Reverse-Primer verwendet, die im Exon 2 und im Intron 2 des *HBB*-Gens hybridisierten. Die Reads wurden mit definierten Referenzsequenzen abgeglichen (Abb. 17, Anhang II und VII).



**Abb. 17: Relative Zusammensetzung der DL- und PL-Mutagenesebanken:** Die Reads der Tiefensequenzierung der DL- und PL-Mutagenesebanken wurden definierten Referenzsequenzen zugeordnet und es folgte die Bestimmung deren relativer Anteile. Die Nomenklatur der Referenzsequenzen beschreibt die Sequenzeigenschaft und die Anzahl der erfolgreich mutierten Basen bzw. folgt einer chronologischen Nummerierung. Beispielsweise charakterisiert die Referenzsequenz *nativ\_8xN* eine korrekte Mutation von 8 bp. Im Fall einer DL-Analyse entspricht dies einer erfolgreich randomisierten Mutagenese. Kryptische Sequenzen enthalten Insertionen bzw. Deletionen außerhalb der Mutationskassette und artifizielle Sequenzen charakterisieren zum Beispiel Primer-Dimere.

Zu 99,1 %  $\pm$  0,2 % konnte die Zusammensetzung der DL- und PL-Mutagenesebanken aufgeklärt werden. Da 66,1 %  $\pm$  0,4 % der Reads ein korrekt mutiertes Motiv (*nativ\_8xN* (DL) und *nativ\_7xN* (PL)) aufwiesen, wurde von einer 34 %igen Synthesefehlerrate ausgegangen. Die unvollständig hydrolysierte Template-DNA (*nativ\_0xN*) der PCR-basierten randomisierten Mutagenesen bildete mit 27,4  $\pm$  5,7% den Hauptanteil der Synthesefehlerrate. Da jedes Motiv mindestens einmal detektiert wurde, konnte von einer ausreichenden Motivabdeckung ausgegangen werden. Anschließend fand die Transfektion dieser voll charakterisierten Mutagenesebanken in die Zelllinien HeLa, HuH7 und SH-SY5Y statt.

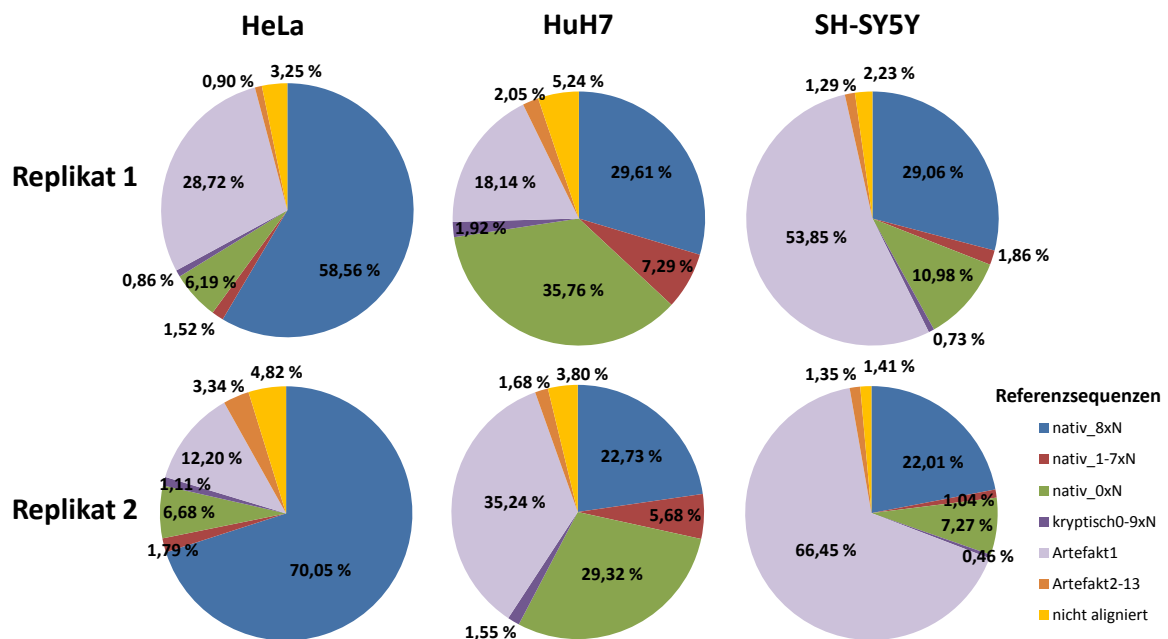
### 5.3.2 Motivabdeckung der transfizierten Mutagenesebanken

Um zu überprüfen, ob die für eine Transfektion von  $1 \times 10^6$  humanen Zellen eingesetzte Menge an Plasmid ausreichte, um eine vollständige Motivabdeckung in den Zellen zu gewährleisten, wurden die systematischen Fehlerquellen, im Speziellen Synthesefehlerrate der Mutagenesebanken und Transfektionseffizienzen, analysiert (Anhang VIII). Dabei wurde ermittelt, dass die

geringste Motivabdeckung  $2,28 \times 10^5$  betrug (Analyse des DL-Reporterbereichs in SH-SY5Y-Zellen). Demnach befanden sich nach der Transfektion ausreichend Plasmide in den Zellen, sodass davon ausgegangen werden konnte, jedes Motiv untersuchen zu können.

### 5.3.3 Sequenzabgleich der Transkript-Pools

Im Anschluss an die Transfektionen der DL- und PL-Mutagenesebanken in die Zelllinien HeLa, HuH7 sowie SH-SY5Y wurden nach 48 h die RNAs isoliert, in cDNAs umgeschrieben und die Transkripte mittels Tiefensequenzierung analysiert. Die verwendeten Forward- und Reverse-Primer hybridisierten im Exon 2 und in der GFP-kodierenden Sequenz des Reporters, sodass ausschließlich die E<sub>ein</sub>-Transkriptisoform sequenziert wurde. Zur Kontrolle der Reproduzierbarkeit der Ergebnisse wurden unabhängige Duplikate (Replik 1 und Replik 2) angefertigt. Im Anhang IX sind die alignierten Bestandteile der DL- und PL-Transkript-Pools aufgelistet. Da die Zusammensetzung des PL-Transkript-Pools der des DLs ähnelt, sind in Abbildung 18 exemplarisch die Bestandteile des DL-Transkript-Pools dargestellt.



**Abb. 18: Relative Zusammensetzung des DL-Transkript-Pools:** Die Reads der Tiefensequenzierung des DL-Transkript-Pools wurden definierten Referenzsequenzen zugeordnet und es erfolgte die Bestimmung deren relativer Anteile. Die Nomenklatur der Referenzsequenzen beschreibt die Sequenzeigenschaft und die Anzahl der mutierten Basen der Reportertranskripte. Die Referenzsequenz *nativ\_8xN* beschreibt eine vollständige Mutation von 8 bp. Kryptische Sequenzen enthalten Insertionen bzw. Deletionen außerhalb der Mutationskassette und artifizielle Sequenzen charakterisieren zum Beispiel Primer-Dimere.

Die marginale Abweichung der zelllinienspezifischen Replikate belegt eine gute Reproduzierbarkeit der Ergebnisse. Insgesamt wurden die Zusammensetzungen der DL- und PL-Transkript-Pools zu  $97,0 \% \pm 1,3 \%$  aufgeklärt. Eine 23x Abdeckung der Motive mit Reads, die eine vollständige DL- bzw. PL-Mutation (*nativ\_8xN* bzw. *nativ\_7xN*) enthielten, stellte die geringste Motivabdeckung dar (Analyse des DL-Reporterbereichs in HuH7-Zellen, Replik 2). Demzufolge

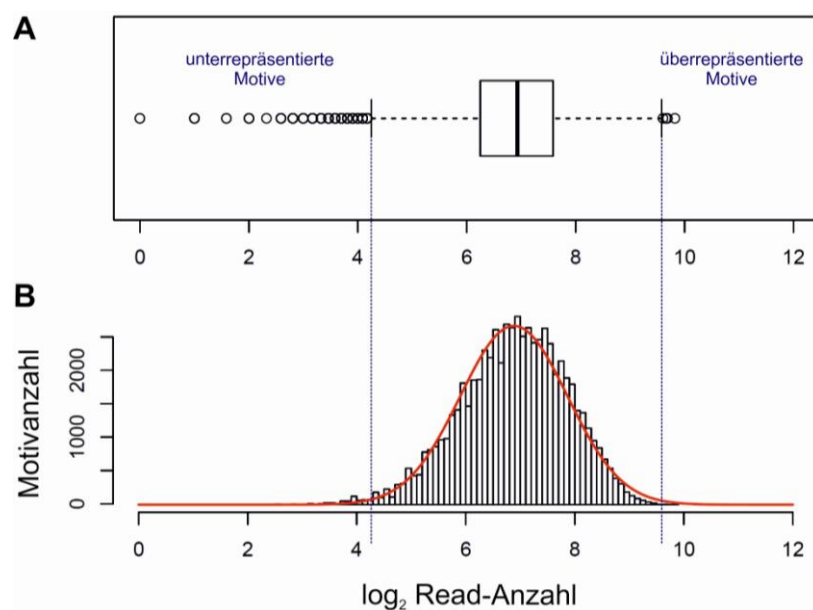
wurden die Transkripte in einer ausreichenden Tiefe sequenziert. Die Referenzsequenz *Artefakt1* beschreibt ein Primer-Dimer, dessen Anteil proportional zur Anzahl an Phasenverschiebungsbasen im Forward-Primer stieg. Diese Basen waren notwendig, um eine Amplikonssequenzierung homogener Sequenzgemische durchzuführen (Abschn. 4.4.6). Während die Transkripte der HeLa-Zellen mit Primern, die maximal eine Phasenverschiebungsbasis beinhalteten, sequenziert wurden, erfolgte die Transkriptanalyse der Zelllinie SH-SY5Y mit Primern, die bis zu sechs Phasenverschiebungsbasen besaßen. Zudem war auffällig, dass die DL- und PL-Transkript-Pools der HuH7-Zellen mit  $30,0\% \pm 6,2\%$  einen deutlich erhöhten  $E_{\text{ein}}$ -Isoformanteil mit vollständig deletierter Mutationskassette (*nativ\_0xN*) beinhalteten, im Gegensatz zu den Zelllinien HeLa mit  $5,7\% \pm 1,0\%$  und SH-SY5Y mit  $7,4\% \pm 2,7\%$ .

### 5.3.4 Häufigkeitsverteilungen der Motive

Zur quantitativen Evaluierung von SRMs erfolgte zunächst die Bestimmung der Häufigkeitsverteilungen der Motive, deren Reads den Referenzsequenzen *nativ\_8xN* (DL-Analyse) und *nativ\_7xN* (PL-Analyse) in den Mutagenesebanken und Transkript-Pools zugewiesen werden konnten. Nach der Verteilung der Daten richteten sich anschließende statistische Analysen.

- **Häufigkeitsverteilungen der Motive in den Mutagenesebanken**

Die Häufigkeitsverteilung der einzelnen Motive in den Mutagenesebanken wurde durch die Berechnung der statistischen Verteilung der Read-Anzahlen pro Motiv bestimmt. Die Datenanalyse ist am Beispiel der DL-Mutagenesebank in Abbildung 19 beschrieben.



**Abb. 19: Häufigkeitsverteilung der Motive in der DL-Mutagenesebank:** **A:** Boxplot zur Darstellung der Read-Verteilung pro Motiv und Identifizierung von unter- und überrepräsentierten Motiven (Ausreißer). Entlang der Abszisse sind die logarithmierten Read-Anzahlen dargestellt. **B:** Häufigkeitsverteilung der logarithmierten Read-Anzahlen (Abszisse) pro Motiv (Ordinate). Die Normalverteilung ist rot dargestellt. Die in A identifizierten über- und unterrepräsentierten Motive sind durch blau gestrichelte Linien abgegrenzt.

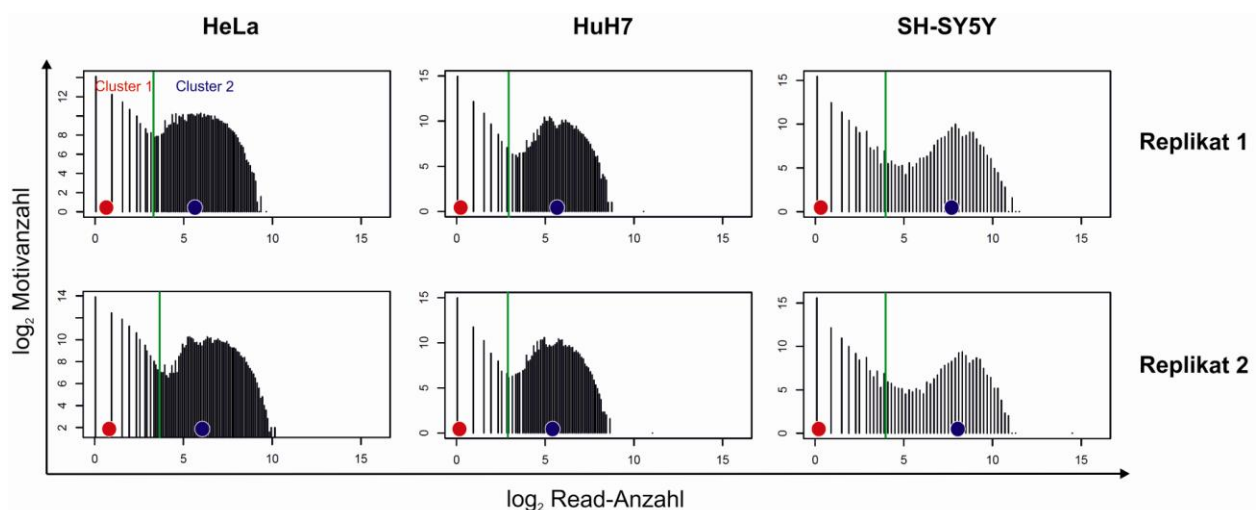
Die graphisch dargestellte Verteilung der Reads und das Ergebnis eines durchgeführten Kolmogorow-Smirnow-Tests (KS-Test) auf Normalverteilung zeigen, dass sich die 9.582.954 Reads der DL-Mutagenesebank gleichmäßig auf die 65536 erwarteten 8-mer Motive verteilen ( $p > 0,05$ ; KS-Test). Die insgesamt 500 identifizierten über- und unterrepräsentierten DL-Motive konnten, nach Rücksprache mit der Firma Microsynth (Balgach, Schweiz), auf einen Primer-Synthesefehler zurückgeführt werden. Diese Motive wurden für weitere Analysen nicht berücksichtigt.

Die Read-Anzahlen pro Motiv der PL-Mutagenesebank waren ebenfalls normalverteilt ( $p > 0,05$ ; KS-Test). Hier wurden 154 Motive als über- und unterrepräsentiert identifiziert und für weitere Analysen nicht verwendet.

### • Häufigkeitsverteilungen der Motive in den Transkript-Pools

Um eine vollständige Datenanalyse der Transkript-Pools im logarithmischen Maßstab zu gewährleisten, wurden zu Beginn der Untersuchung nicht detektierte Motive mit einem Pseudo-Read-Count von 1 versehen.

Wie in Abbildung 20 dargestellt und durch einen KS-Test berechnet ( $p < 0,05$ ; KS-Test), wiesen die Reads pro Motiv des DL-Transkript-Pools keine Normalverteilung auf, sondern waren bimodal verteilt. Mit einer Clusterzentrenanalyse, durchgeführt mittels *k*-Means-Algorithmus (Distanzfunktion), wurden die Zentren der zwei Cluster identifiziert und jedem Cluster die entsprechenden Motive zugeordnet. Der PL-Transkript-Pool wies eine nahezu identische Verteilung der Reads pro Motiv auf ( $p < 0,05$ ; KS-Test) (Anhang X).



**Abb. 20: Häufigkeitsverteilung der Motive in den DL-Transkript-Pools:** Darstellung der Häufigkeitsverteilungen logarithmierter Read-Anzahlen (Abszisse) und korrespondierender logarithmierter Motivanzahlen (Ordinate). Die roten und blauen Punkte symbolisieren die Clusterzentren. Die grünen Linien markieren die Clustergrenzen.

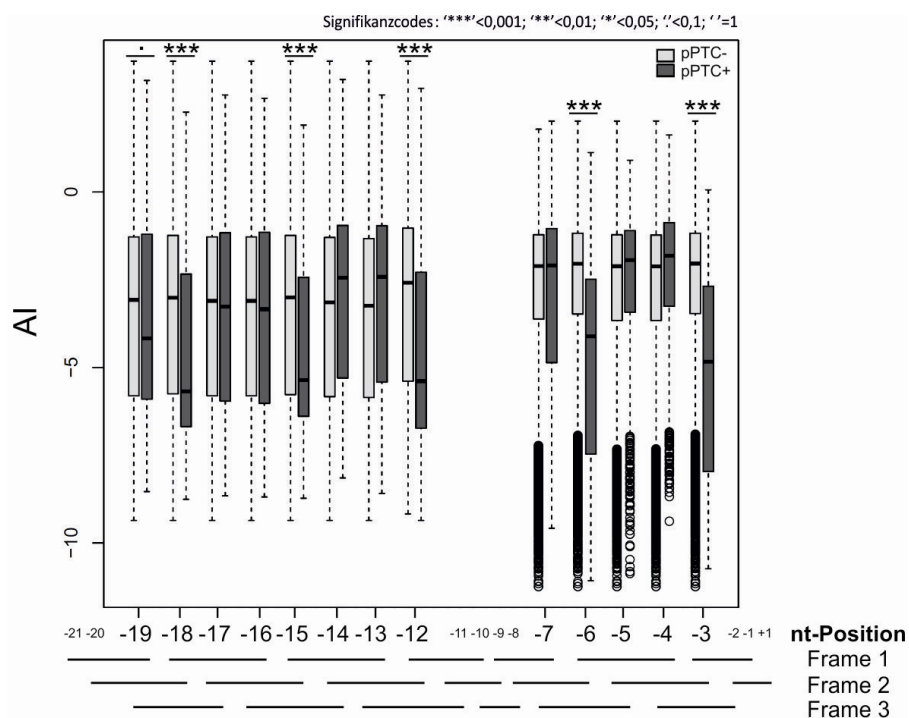
In HeLa-Zellen erzeugten  $92,1 \% \pm 6,7 \%$ , in HuH7-Zellen  $69,8 \% \pm 9,5 \%$  und in SH-SY5Y-Zellen  $53,7 \% \pm 9,5 \%$  der erwarteten DL- und PL-Motive eine konstitutive Spleißreaktion ( $E_{\text{ein}}$ ). Aufgrund dessen, dass die meisten Motive eine gemeinsame Read-Anzahl von 0 aufwiesen,



Silencer jedoch in den seltensten Fällen zu einer 100 %igen Spleißinhibition führen, wurde die Hypothese aufgestellt, dass die bimodale Motivverteilung durch weitere Effekte entstand.

### 5.3.5 Aufklärung des Mischeffektes in den Transkriptdaten

Um SRMs quantitativ evaluieren zu können, musste der Mischeffekt in den Transkriptdaten identifiziert und mathematisch bereinigt werden. Der NMD-Mechanismus (Abschn. 3.4.2) war der einzig nachweisbare Effekt, der zur Reduktion der Motivdiversität im Transkript-Pool hätte beitragen können. Zur Analyse des NMD-Effektes wurde für jedes Motiv ein Anreicherungsindex (AI) bestimmt. Der AI eines Motivs berechnete sich aus der Differenz der logarithmierten ( $\log_2$ ) Read-Anzahl des Transkript-Pools und der zugehörigen Mutagenesebank. Zusätzlich wurden die putativen vorzeitigen Translationsstoppkodonen (engl. *putative premature termination codons*, pPTCs) TAA, TAG und TGA in den DL- und PL-Motiven identifiziert. Ein positionsabhängiger Vergleich der AI-Verteilungen von Motiven mit vorhandenem (pPTC+) und nicht vorhandenem pPTC (pPTC-) wurde mittels einseitigen Wilcoxon-Mann-Whitney-Tests (U-Test) durchgeführt. Ein signifikanter Unterschied der AI-Verteilungen war ein mathematischer Beweis für einen NMD-Effekt (Abb. 21 und Anhang XI).



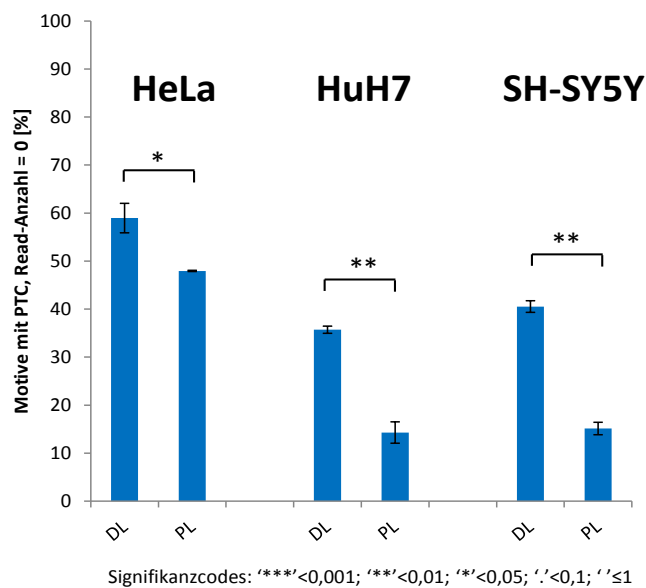
**Abb. 21: Identifizierung eines NMD-Effektes im DL- und PL-Transkript-Pool in HeLa-Zellen (Replik 1):** Graphische Darstellung der Verteilungen der Anreicherungsindizes (AI) (Ordinate) der Motive ohne (pPTC-) und mit putativem vorzeitigem Translationsstoppkodon (pPTC+) (Abszisse) in Abhängigkeit der Nukleotidpositionen (nt-Positionen). Frame 1 charakterisiert den nativen *HBB*-Leserahmen. Der Vergleich der AI-Verteilungen der positionsabhängigen pPTC+ und pPTC- wurde mittels einseitigem U-Test durchgeführt.

In den Zelllinien HeLa, HuH7 und SH-SY5Y konnte herausgefunden werden, dass ein PTC im nativen *HBB*-Leserahmen, beginnend an den Positionen -18, -15, -12, -6 und -3 des Introns 2



eine hoch signifikante Reduktion des Als ( $p < 0,001$ ; U-Test) zur Folge hatte. Dies war der Beweis eines NMD-Effektes in den Transkriptdaten. Durch die Identifizierung des NMD-Effektes konnte ein Faktor, der zur Entstehung der bimodalen Motivverteilung in den Transkript-Pools beitrug, bestimmt werden.

Um zu überprüfen, ob es in Abhängigkeit von Locus und Zelltyp unterschiedlich starke NMD-Effekte gab, wurden die Transkripte, deren Read-Anzahl bei 0 lag und die zudem einen PTC aufwiesen, gesondert betrachtet. Hierfür wurde die relative Motivanzahl der besagten Transkripte aus den Replikaten 1 und 2 ermittelt und in Abhängigkeit der DL- und PL-Analysen mittels zweiseitigen t-Tests nach Student verglichen (Abb. 22 und Anhang XII).



**Abb. 22: Quantifizierung des NMD-Effektes in Abhängigkeit von Locus und Zelltyp:** Die Ordinate stellt den Anteil der Motive dar, die ein vorzeitiges Translationsstoppkodon (PTC) enthielten und nicht im Transkript-Pool detektiert wurden. Der positionsabhängige Vergleich wurde mittels zweiseitigen t-Tests nach Student berechnet.

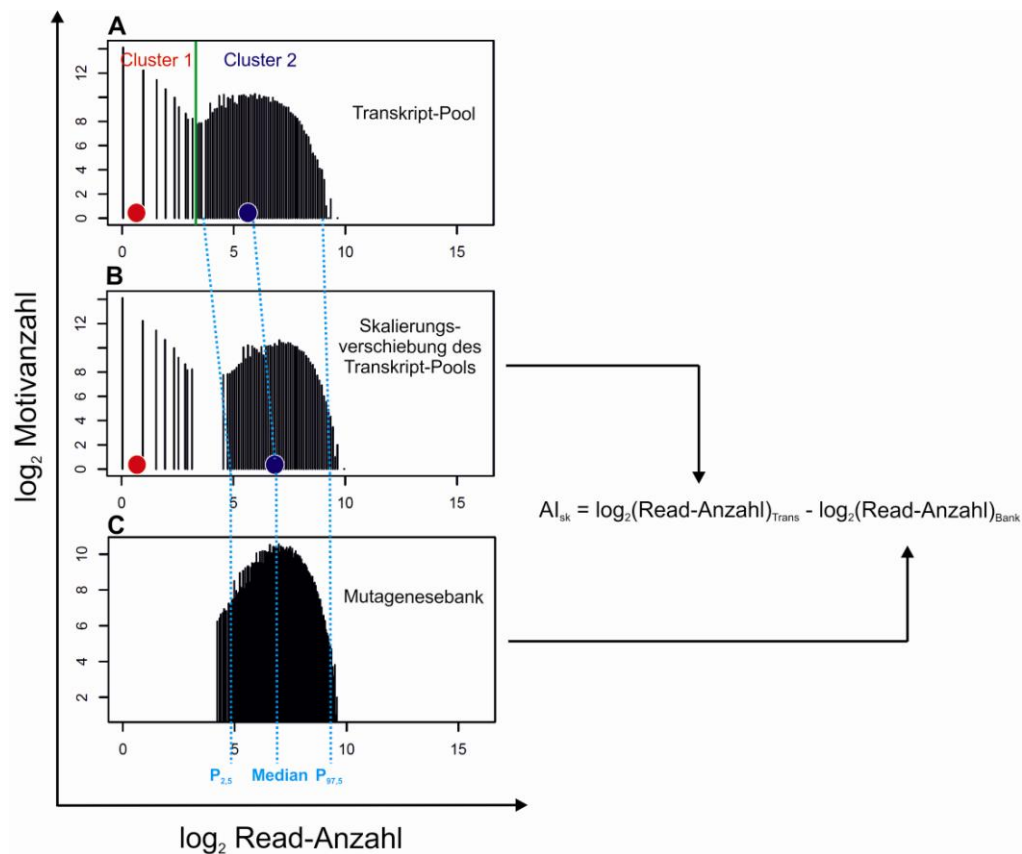
Die Berechnungen zeigen, dass mit zunehmendem Abstand des PTCs distal zur Exon-Exon-Grenze ein signifikant stärkerer NMD-Effekt besteht, unabhängig vom Zelltyp. Dabei wiesen HeLa-Zellen den effektivsten NMD-vermittelten mRNA-Abbau auf, während dieser in der Zelllinie HuH7 am schwächsten ausgeprägt war. Mit einem Unterschied von  $25,4 \% \pm 2,5 \%$  zwischen den DL- und PL-Motiven zeigten die SH-SY5Y-Zellen den stärksten positionsabhängigen Effekt.

### 5.3.6 Berechnung des skalierten Anreicherungsindex ( $AI_{sk}$ )

Zur Bestimmung und zum Vergleich des Einflusses von SRMs auf die Spleißaktivität einer 5'-Spleißstellenerkennung in verschiedenen Zelllinien, wurde ein skaliertes Anreicherungsindex ( $AI_{sk}$ ) gebildet. Hierfür wurde zunächst das Skalenniveau des Transkript-Pools der Mutagenesebank angepasst. Der Median, das Perzentil 2,5 und das Perzentil 97,5 des Transkript-Pools wurden mit den korrespondierenden Werten der Mutagenesebank abgeglichen. So verteilten

sich 95 % der Transkriptdaten auf dem Skalenniveau der Mutagenesebank (Ke et al. 2011). Die Verschiebung der Skalierung wurde lediglich mit den Transkriptdaten des Clusters 2 (Abschn. 5.3.4) durchgeführt, da das Cluster 1 (Abschn. 5.3.5) zusätzliche Effekte beinhaltet und eine Effektvermischung während der Identifizierung von SRMs vermieden werden sollte. Im Anschluss wurde für jedes Motiv die Differenz der Read-Anzahlen von Transkript-Pool und zugehöriger Mutagenesebank gebildet.

Einen Vorteil dieser Skalierung stellte die direkte Einschätzung des Einflusses der Motive auf die Spleißaktivität dar. War der  $AI_{sk}$  größer als 0, wurde von einer Anreicherung der  $E_{ein}$ -Isoform ausgegangen. Ein negativer  $AI_{sk}$  charakterisierte eine Abreicherung der  $E_{ein}$ -Isoform. In Abbildung 23 ist die Berechnung des  $AI_{sk}$  exemplarisch am DL-Reporterbereich in HeLa-Zellen (Replik 1) abgebildet.



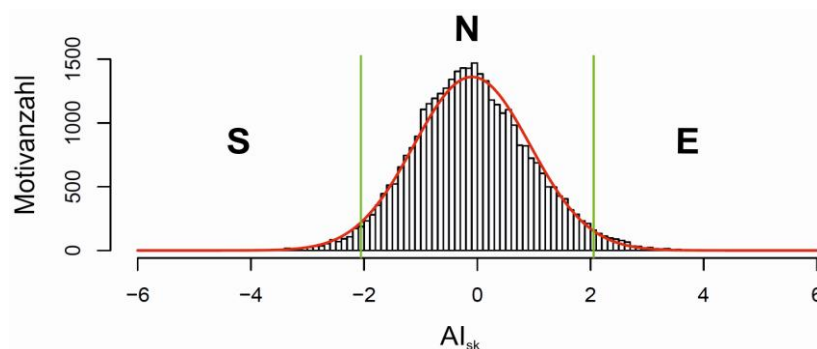
**Abb. 23: Schematische Darstellung der Berechnung des  $AI_{sk}$  des DL-Reporterbereichs in HeLa-Zellen (Replik 1):** **A:** Die grüne Linie markiert die Clustergrenze des Transkript-Pools, ab der die Skalierungsverschiebung des Clusters 2 ansetzt. Die roten und blauen Punkte markieren die Clusterzentren. **B:** Der Abgleich des Medians, des Perzentils 2,5 ( $P_{2,5}$ ) und des Perzentils 97,5 ( $P_{97,5}$ ) des Clusters 2 mit den korrespondierenden Werten der Mutagenesebank vermittelt eine Stauchung der Read-Verteilung des Transkript-Pools. **C:** Mutagenesebank, die zum Skalierungsabgleich verwendet wurde. Der  $AI_{sk}$  eines Motivs berechnet sich aus der Differenz der logarithmierten Read-Anzahl des neu skalierten Clusters 2 (Trans) und der logarithmierten Read-Anzahl der Mutagenesebank (Bank).

### 5.3.7 Identifizierung von Enhancer- und Silencer-Sequenzen

Zur ausschließlichen Identifizierung von SRMs, die in verstärkender (Enhancer) bzw. hemmender Weise (Silencer) die Spleißaktivität beeinflussten, wurden zunächst Motive, die einen PTC

beinhalteten, aus der Analyse entfernt. Anschließend sollte der Mittelwert der  $AI_{sk}$  der Replikate einer Zelllinie gebildet werden. Hierfür musste vorab bestätigt werden, dass die Replikate einer Zelllinie zueinander ähnlicher sind, als die Replikate verschiedener Zelllinien. Daher erfolgte die Bestimmung des Korrelationskoeffizienten ( $R^2$ ) jeder möglichen Replikatkombination nach Pearson (Anhang XIII). Die Korrelationsmessung zeigt, dass die zelllinienspezifischen Replikate mit einem  $R^2$  von  $0,224 \pm 0,05$  eine moderate Korrelation aufwiesen. Replikate unterschiedlichen Zelltyps besaßen mit einem  $R^2$  von  $0,048 \pm 0,04$  keine Gemeinsamkeiten. Aufgrund dessen konnte der Mittelwert der  $AI_{sk}$  der Replikate einer Zelllinie gebildet werden.

Für eine Identifizierung von Enhancern und Silencern wurde die Häufigkeitsverteilung der durchschnittlichen  $AI_{sk}$  in Abhängigkeit der Motive berechnet. Aufgrund der Normalverteilung der  $AI_{sk}$  ( $p > 0,05$ ; KS-Test) konnten durch Einführung einer Signifikanzschwelle ( $\pm 2x$  Standardabweichung der Häufigkeitsverteilung) Enhancer- und Silencer-Sequenzen bestimmt werden. Die Signifikanzschwelle gab mit minimaler Messunsicherheit den  $AI_{sk}$  an, ab dem der quantifizierte Einfluss eines Motivs auf die Spleißaktivität signifikant war. Dementsprechend wurden Motive mit einem  $AI_{sk}$  größer als die  $2x$  Standardabweichung als Enhancer, und Motive, deren  $AI_{sk}$  kleiner als die  $2x$  Standardabweichung war, als Silencer identifiziert (Anhang XIV). In Abbildung 24 ist die Analyse am Beispiel des DL-Reporterbereichs in HeLa-Zellen dargestellt.



**Abb. 24: Identifizierte Enhancer- und Silencer-Sequenzen des DL-Reporterbereichs in HeLa-Zellen:** Graphische Darstellung der  $AI_{sk}$ -Verteilung (Abszisse) in Abhängigkeit der Motivanzahl (Ordinate). Die Normalverteilung ist rot dargestellt. Die grünen Linien symbolisieren die Signifikanzschwellen ( $\pm 2x$  Standardabweichung) und unterteilen die Motive in Silencer (S)-, neutrale (N) und Enhancer-Sequenzen (E).

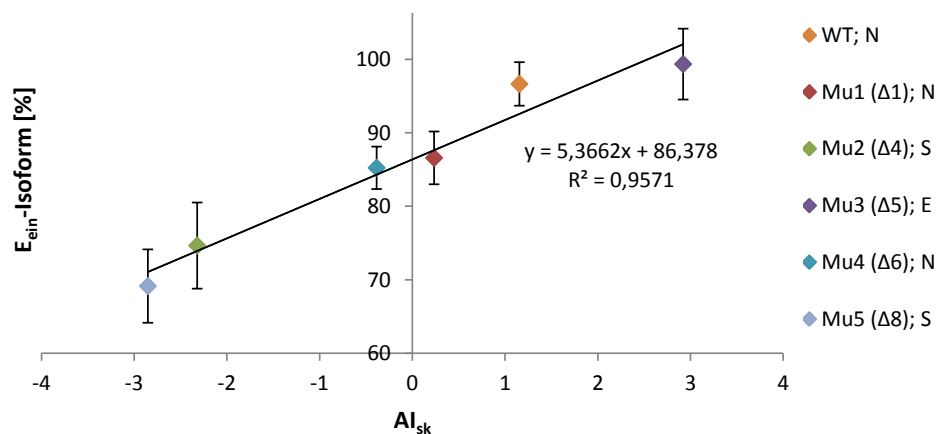
In Abhängigkeit von Locus und Zelltyp sind in Tabelle 5 die Anzahlen der identifizierten Enhancer- und Silencer-Sequenzen aufgelistet. Die Sequenzen und die zugehörigen  $AI_{sk}$  der Enhancer und Silencer befinden sich im Anhang XIV.

**Tab. 5: Anzahlen identifizierter Enhancer- und Silencer-Sequenzen in Abhängigkeit von Locus und Zelltyp**

Motivanzahlen	HeLa		HuH7		SH-SY5Y	
	DL	PL	DL	PL	DL	PL
Gesamt-Motivanzahl	34988	13807	27316	8945	9796	4926
• Enhancer	816	353	419	116	121	70
• neutrale Sequenzen	33347	13099	25979	8527	9286	4688
• Silencer	825	355	918	302	389	168

## 5.4 Validierung der identifizierten SRMs

Um zu überprüfen, ob der quantitativ evaluierte Einfluss ( $AI_{sk}$ ) der identifizierten SRMs auf eine 5'-Spleißstellenerkennung reproduzierbar ist und demnach Gültigkeit besitzt, wurden fünf zufällig ausgewählte Motive des DL-Reporterbereichs mit ortsgerichteten Mutagenesen rekonstruiert, die rHBBg-Plasmide in die Zelllinien HeLa, HuH7 und SH-SY5Y transfiziert und die  $E_{ein}$ -Isoformen mittels FACS-Analysen quantifiziert. Die quantifizierten  $E_{ein}$ -Isoformen wurden im Anschluss mit den spezifischen  $AI_{sk}$  verglichen (Abb. 25 und Anhang XV). Es wurde die Hypothese aufgestellt, dass Enhancer-Sequenzen mit einem hohen  $AI_{sk}$  zu einer verstärkten  $E_{ein}$ -Reaktion führen und demnach mit einem hohen  $E_{ein}$ -Isoformanteil korrelieren. Im Gegensatz dazu sollten Silencer-Sequenzen mit einem niedrigen  $AI_{sk}$  einen niedrigen  $E_{ein}$ -Isoformanteil vermitteln.



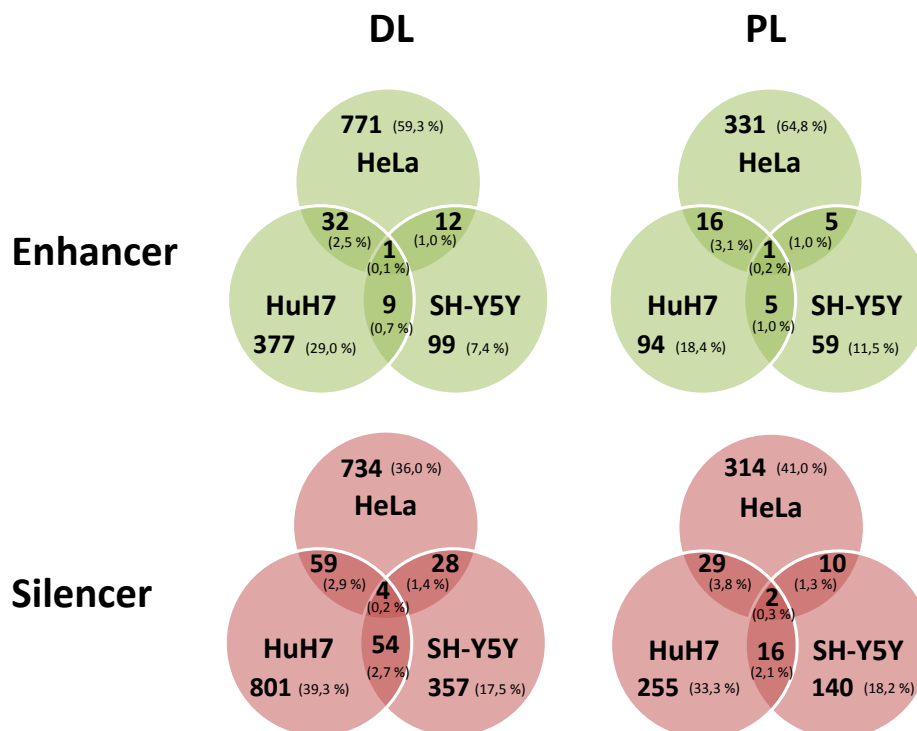
**Abb. 25: Validierung ausgewählter SRMs des DL-Reporterbereichs in HeLa-Zellen:** Entlang der Abszisse sind die  $AI_{sk}$  des Wildtypmotivs und fünf verschiedener DL-Motive (Mu1-5) abgetragen. „Δ“ symbolisiert die Anzahl an Basen, in denen sich die DL-Motive vom Wildtypmotiv unterscheiden. Die Motive wurden mithilfe des  $AI_{sk}$  in Enhancer- (E), neutrale (N) und Silencer-Sequenzen (S) unterteilt. Die Ordinate repräsentiert die mittels FACS ermittelten Anteile an konstitutiv gespleißten Isoformen ( $E_{ein}$ ).

Die Ergebnisse zeigen, dass die  $AI_{sk}$  und die zusätzlich quantifizierten  $E_{ein}$ -Isoformen stark positiv korrelierten ( $R^2 \geq 0,9$ ) und ein linearer Zusammenhang bestand. Aufgrund dessen konnte erfolgreich der quantitativ evaluierte Einfluss auf die Spleißaktivität reproduziert und die Gültigkeit der identifizierten SRMs bestätigt werden.

## 5.5 Charakterisierung der identifizierten SRMs

### 5.5.1 Zelllinienübergreifender Vergleich der SRMs

Zur Bestimmung des Einflusses des Zelltyps auf das Spleißen wurden die in den Zelllinien HeLa, HuH7 und SH-SY5Y identifizierten Enhancer- und Silencer-Sequenzen der DL- und PL-Reporterbereiche zelllinienübergreifend miteinander verglichen (Abb. 26).



**Abb. 26: Schnittmengendiagramme des zelllinienübergreifenden Vergleichs der SRMs:** Die Schnittmengendiagramme stellen die absoluten (prozentualen) Motivanzahlen in Abhängigkeit der Zelllinienübereinstimmungen dar.

Die gemeinsame Schnittmenge an Enhancern und Silencern in den drei Zelllinien betrug insgesamt  $0,18 \% \pm 0,07$ , wobei die Zelllinien HeLa und HuH7 mit  $3,1 \% \pm 0,5$  % Übereinstimmung die größten Zellliniengemeinsamkeiten aufwiesen. In Tabelle 6 sind die gemeinsamen Motive der drei Zelllinien in Abhängigkeit vom Locus aufgelistet.

**Tab. 6: Gemeinsame DL- und PL-Motive der Zelllinien HeLa, HuH7 und SH-SY5Y**

spleißregulatorische Motive	DL (-19 bis -12)	PL (-7 bis -1)
Enhancer-Sequenzen	TCGCTATA	TTCGCC
Silencer-Sequenzen	CGCGGCGG	ATTGGGT
	CTTCGTGC	GGGAGGT
	GGGGGTTCG	
	GTTGGGTA	

Durch einen Abgleich der Enhancer- und Silencer-Sequenzen mithilfe des Programms *ClustalW* (EMBL) konnten keine Sequenzhomologien festgestellt werden.

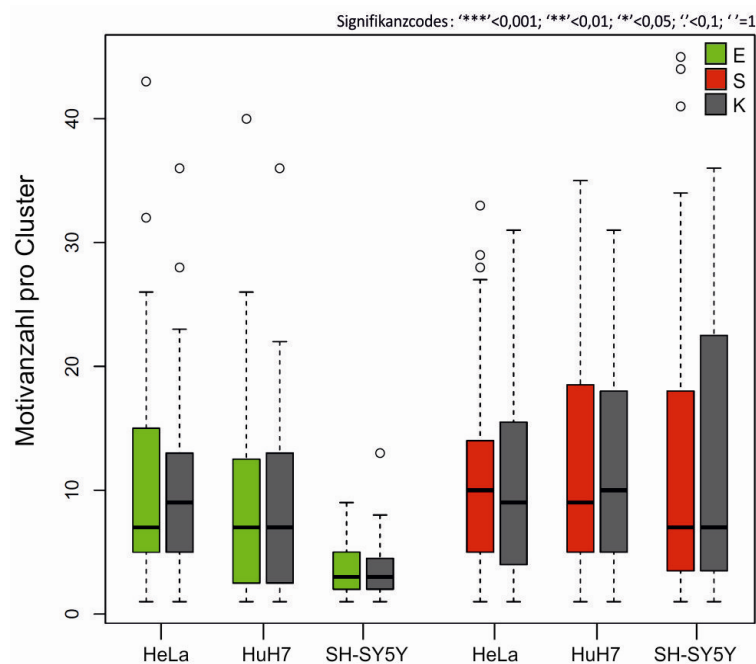
Da die drei untersuchten Zelllinien nahezu keine gemeinsamen Enhancer- und Silencer-Sequenzen besaßen, kann davon ausgegangen werden, dass der Einfluss der identifizierten SRMs auf eine 5'-Spleißstellenerkennung stark vom Zelltyp geprägt war.

Letztendlich konnten unabhängig von der Zelllinie insgesamt 1812 unterschiedliche Enhancer-Sequenzen identifiziert werden, davon 1301 Sequenzen distal (DL) und 511 Sequenzen proximal (PL) der 5'-Spleißstelle. Von den insgesamt 2803 Silencer-Sequenzen wurden 2037 Sequenzen distal (DL) und 766 Sequenzen proximal (PL) der 5'-Spleißstelle identifiziert.

### 5.5.2 Sequenzhomologien von SRMs

Die Bestimmung von Sequenzhomologien von Enhancer- bzw. Silencer-Sequenzen wurde durchgeführt, um identische Sequenzabschnitte zu identifizieren und auf dieser Grundlage die SRMs zu gruppieren.

Es erfolgte eine Sequenz-Clusteranalyse mit dem Programm CD-HIT-EST, welches jedes analysierte SRM einem definierten Cluster entsprechend der Sequenzhomologie zuwies (Huang et al. 2010). Die Homologiegrenze lag bei 85 %. Betrachtet wurde die vollständige Motivkomplexität der DL- und PL-Reporterbereiche (Abschn. 5.1.3). Ob die gebildeten Cluster Gültigkeit besaßen, wurde durch einen Vergleich mit einem Kontrollmodell getestet. Die Erstellung des Kontrollmodells erfolgte mithilfe des Programms *List Shuffler* (<<http://www.amdz.com/shuffler.php>>). Hierfür wurden zufällig Enhancer-, neutrale und Silencer-Sequenzen, denen definitiv keine Sequenzhomologien zugrunde lagen, ausgewählt und als Kontrolle eingesetzt. Aus dem Vergleich mit dem Kontrollmodell sollte sich ergeben, dass eine Clusteranalyse der Enhancer- bzw. Silencer-Sequenzen signifikant mehr Sequenzhomologien aufweist und demnach weniger Cluster mit größeren Motivanzahlen bildet. Jede Clusteranalyse der SRMs wurde mit 100 verschiedenen Kontrollmodellen verglichen (zweiseitiger U-Test), sodass der resultierende Vertrauensbereich der Analyse 100 % betrug (Abb. 27 und Anhang XVI).

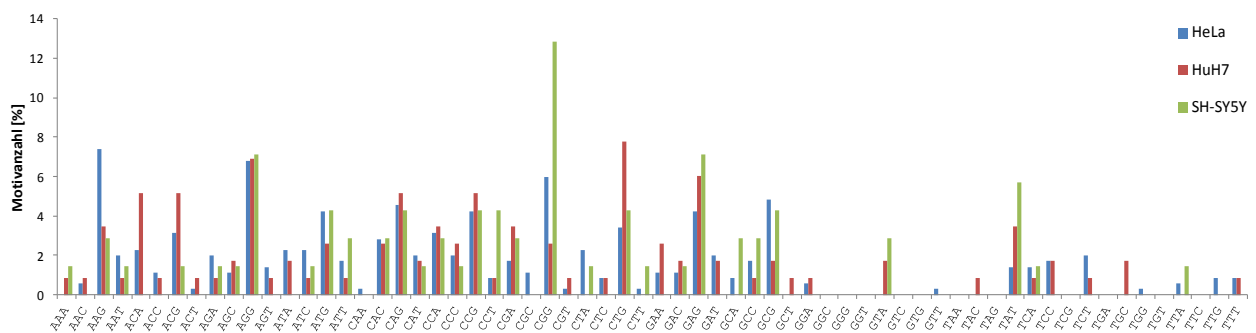


**Abb. 27: Clusteranalyse der SRMs des DL-Reporterbereichs:** Graphische Darstellung der den identifizierten Clustern zugewiesenen Motivanzahlen (Ordinate) der Enhancer- (E), Silencer- (S) und Kontrollanalysen (K) in Abhängigkeit vom Zelltyp (Abszisse). Die Motivanzahlen pro Enhancer- bzw. Silencer-Cluster weisen keinen signifikanten Unterschied zum jeweiligen Kontrollmodell auf ( $p > 0,05$ ; zweiseitiger U-Test).

Mithilfe der ausgeführten Clusteranalyse konnten keine Sequenzhomologien innerhalb der Enhancer- bzw. Silencer-Sequenzen bestimmt werden.

### 5.5.3 Analyse der Bindestelle der Spleißosomkomponente U1-snRNP

Die Positionen -3 bis -1 des PL-Reporterbereichs bildeten den exonischen Anteil der Bindestelle der Spleißosomkomponente U1-snRNP (Abschn. 3.2.1-2). Um Aussagen über die Sequenz einer starken Bindestelle machen zu können, wurden diese drei Basen in Abhängigkeit vom Zelltyp gesondert betrachtet. In diesem Zusammenhang wurden die Enhancer-Sequenzen des PL-Reporterbereichs, die ein starkes 5'-Spleißstellensignal erzeugten, entsprechend ihrer Positionen -3 bis -1 einem von  $4^3$  möglichen Triplets zugeordnet und deren relative Anzahl bestimmt (Abb. 28).



**Abb. 28: Analyse der Bindestelle des U1-snRNPs:** Die Ordinate repräsentiert den prozentualen Anteil der Triplets in den Enhancer-Sequenzen an den Positionen -3 bis -1 des PL-Reporterbereichs, die den 64 Kombinationsmöglichkeiten (Abszisse) zugeordnet wurden.

Die Analyse zeigt, dass die Triplets AGG mit 7,4 % in HeLa-Zellen, CTG mit 7,6 % in HuH7-Zellen und CGG mit 12,9 % in SH-SY5Y-Zellen im stromaufwärts flankierenden Bereich der 5'-Spleißstelle (Position -3 bis -1 des PL-Reporterbereichs) das Spleißsignal am effektivsten vermittelten. Die Triplets GGC, GGG, GGT, GTC, GTG, TAA, TAG, TGA, TGT und TTC konnten unabhängig vom Zelltyp nicht detektiert werden, da diese den NMD-Mechanismus aktivierten (TAA, TAG, TGA) bzw. aufgrund des erhöhten GT-Anteils vermutlich einen alternativen Spleißdonor beinhalteten.

### 5.5.4 Annotation der SRMs

Die Annotation der SRMs beinhaltete die funktionelle Zuordnung bzw. das Erstellen von Assoziationen zwischen Enhancer/Silencer-Sequenzen und bereits mathematisch beschriebenen, spleißregulatorischen Faktoren. Auf diese Weise konnten beispielsweise Bindeproteine der identifizierten SRMs ermittelt werden.

Für eine funktionelle Motivannotation wurden zunächst der GC-Gehalt, der Purin-Gehalt und die Abweichung des zu analysierenden Motivs vom Wildtypmotiv (WA) bestimmt. Zur Aufklärung der Sekundärstrukturanteile der SRMs wurde der *EF-Score* berechnet (Hiller, Pudimat, et al. 2006). Der *EF-Score* ist ein Maß für die Entropie der prä-mRNA. Je größer die Entropie und damit der Anteil an Einzelstrangregionen in einer prä-mRNA ist, desto größer ist der *EF-Score*.

Mithilfe des Vorhersage-Programms *SFmap* wurden zudem RNA-Bindepartner der Enhancer- und Silencer-Sequenzen identifiziert (Akerman et al. 2009). Dabei wurde die größtmögliche Sequenzhomologie zwischen den identifizierten SRMs und den bereits charakterisierten RNA-Motiven, deren Bindepartner bekannt sind, ermittelt und als *S-Score* quantifiziert.

Die quantifizierten spleißregulatorischen Faktoren der Enhancer- und Silencer-Sequenzen wurden auf Normalverteilung getestet ( $p > 0,017$ ; KS-Test mit Bonferroni-korrigiertem Signifikanzniveau) und mittels Zwei-Stichproben-t-Tests für unabhängige Stichproben miteinander verglichen. In Abhängigkeit vom t-Test konnten im Anschluss die Effektgröße ( $d$ ) und die Effektkorrelation ( $r$ ) bestimmt werden (Abschn. 4.6). Wies ein signifikanter t-Test eine absolute Effektgröße  $|d| > 0,5$  auf, wurde die Assoziation zwischen SRM und Spleißfaktor nach Cohens Standard als stark eingestuft (Cohen 1992). Das Vorzeichen der Effektgröße veranschaulichte dabei die Richtung des Effektes. Während positive Effektgrößen mit Enhancer-Effekten assoziiert waren, wurden negative Effektgrößen Silencer-Effekten zugeordnet. Zum Vergleich der Effektstärke der einzelnen Motivannotationen wurde die Effektkorrelation  $r$  berechnet, deren Maximum bei 1 und deren Minimum bei 0 lag. In Abbildung 29 und im Anhang XVII sind die Ergebnisse der Assoziationsstudie aufgezeigt.



**Abb. 29: Annotation der Enhancer- und Silencer-Sequenzen:** Das Skalenniveau stellt die berechneten Effektkorrelationen zwischen den DL/PL-Motiven und den spleißregulatorischen Faktoren dar, mit einem Minimalwert von 0 und einem Maximalwert von 1. Dabei wird zwischen Enhancer- und Silencer-Effekten differenziert.



Trotz dessen, dass die identifizierten Enhancer- bzw. Silencer-Sequenzen keine Sequenzhomologien aufwiesen, konnte erstaunlicher Weise nahezu immer das gleiche Set an spleißregulatorischen Faktoren ermittelt werden, die mit den SRMs assoziiert waren. Dabei wiesen die Spleißfaktoren YB1 mit  $0,62 \pm 0,08$  und hnRNP H/F mit  $0,57 \pm 0,17$  unabhängig von Locus und Zelltyp die größten Effektkorrelationen auf. Während YB1 als Spleißaktivator identifiziert werden konnte, wurde hnRNP H/F eine Spleißrepressorfunktion zugeordnet. Nach Cohens Standard stellen Effektkorrelationen über 0,5 einen sehr starken Effekt dar. Ebenfalls konnten die Einflüsse der RNA-Sekundärstruktur mit  $0,44 \pm 0,07$  und des GC-Gehaltes der SRMs mit  $0,53 \pm 0,14$  auf die Erkennung einer 5'-Spleißstelle nachgewiesen werden.

Inwieweit die identifizierten spleißregulatorischen Faktoren den quantifizierten Einfluss ( $AI_{sk}$ ) auf die Spleißaktivität beschrieben, wurde mittels linearer Regression ermittelt (Anhang XVIII). Als abhängige Variable der Regressionsmodelle wurde der  $AI_{sk}$  der Enhancer-, neutralen und Silencer-Sequenzen verwendet (Abschn. 5.3.7). Die ermittelten Bestimmtheitsmaße ( $R^2$ ) dienten zur Abschätzung der Varianzaufklärung des  $AI_{sk}$  durch die bekannten spleißregulatorischen Faktoren. Insgesamt konnten maximal 32,4 % (Analyse des DL-Reporterbereichs in SH-SY5Y-Zellen) der Varianz des  $AI_{sk}$  aufgeklärt werden. Das durchschnittliche Bestimmtheitsmaß betrug 20,1 %. Demzufolge konnte mithilfe der Motivannotationen in einer moderaten Komplexität der quantitativ evaluierte Einfluss der identifizierten SRMs auf die Spleißaktivität einer 5'-Spleißstellenerkennung beschrieben werden.

## 6 DISKUSSION

Im Rahmen dieser Arbeit wurden systematisch spleißregulatorische Motive (SRMs) untersucht, die essentiell für die 5'-Spleißstellenerkennung sind, die für die Assemblierung des Spleißosoms relevant sind und welche die Untersuchung von latenten 5'-Spleißstellen ermöglichen. Zudem sollten die identifizierten Motive die Vorhersage der Spleißrelevanz von Mutationen im humanen Genom vereinfachen, da diese durch *de novo* Generierung funktioneller oder fehlerhafter SRMs entscheidenden Einfluss auf die Spleißaktivität haben können. Um die Ziele zu erreichen, wurde zunächst eine Methode etabliert, mit der SRMs kontext- und zelltypabhängig identifiziert werden konnten.

### 6.1 Etablierte Methode zur quantitativen Evaluierung von SRMs

Die im Zuge dieser Arbeit etablierte Methode zur quantitativen Evaluierung von SRMs erfolgte auf der Basis eines Spleißreportersystems. Das hierfür konstruierte rHBBg-Plasmid beinhaltet das komplette *HBB*-Gen (Exon 1-3), ohne die nativen 5'- und 3'-UTRs, flankiert von zwei Fluoreszenzprotein-kodierenden Genen.

Im Gegensatz zu dem etablierten Spleißreportersystem basierten bisherige SRM-Studien lediglich auf der Analyse von unvollständigen Genkontexten (Barash et al. 2010; Ke et al. 2011). Zum Beispiel identifizierte Wang et al. 2004 ESSs mithilfe eines artifiziellen Spleißreportersystems, dessen Exon 1 und 3 die kodierende Sequenz des EGFP-Proteins (engl. *enhanced GFP*) umfasste und dessen Exon 2 das des *DHFR*-Gens (Dihydrofolatreduktase des chinesischen Hamsters) mit angrenzenden Introns war (Wang et al. 2004). Welche biologische Relevanz jedoch der vollständige Genkontext auf eine Spleißreaktion besitzt, veranschaulicht der SNP rs6438552 (C>T) des Gens *GSK3B* (Glykogensynthase-Kinase-3 $\beta$ ). Dieser SNP ist in einem SRM im Intron 5 lokalisiert und initiiert den Ausschluss von Exon 9 stromabwärts über drei Exons hinweg. Durch den Verlust des SRMs wird ein verkürztes Protein synthetisiert, welches zur Phosphorylierung des Tau-Proteins beiträgt und mit der Pathogenese der Parkinson-Krankheit assoziiert ist (Kwok et al. 2005). Da durch die bislang verwendeten Spleißreportersysteme nie der gesamte Genkontext betrachtet worden ist, stellt das rHBBg-Reportersystem ein neuartiges Modell dar. Zudem ermöglicht es die Detektion von konstitutiven und alternativen Spleißreaktionen mittels fluoreszenzbasierter Analysemethoden. So wird im Fall einer konstitutiven Spleißreaktion ein dual fluoreszierendes Protein und im Fall eines alternativen Spleißereignisses ein Protein mit nur einer fluoreszierenden Domäne synthetisiert. Bei ausbleibender Fluoreszenz lassen sich sogar Rückschlüsse auf einen gezielten mRNA-Abbau (engl. *nonsense-mediated mRNA decay*, NMD) ziehen. Die mittels FACS-Analysen quantifizierten Spleißvarianten konnten ebenfalls auf Transkriptebene bestätigt werden, sodass schnell und

kostengünstig das Verhältnis der Transkriptisoformen bestimmt werden kann. Darin liegt ein wesentlicher Unterschied zu den bislang verwendeten Spleißreportersystemen, die in der Regel mit nur einer Fluoreszenz arbeiten (Singh & Cooper 2006; Schoft et al. 2007; Zhang et al. 2009; Somarelli et al. 2013).

Bioinformatisch und experimentell konnten mit dem Reportersystem spleißregulatorische Bereiche bis zu 20 bp stromaufwärts des *HBB*-Introns 2 (L<sub>-19bis-12</sub> (DL) und L<sub>-7bis-1</sub> (PL)) identifiziert werden. Damit bestätigte sich die Hypothese, dass SRMs eine direkte Nachbarschaft zur Spleißstelle aufweisen (Rosenow et al. 2002; Willie & Majewski 2004). Die identifizierten Loci wurden in ausreichender Komplexität und mit einer Syntheseeffizienz von  $66,1 \% \pm 0,4 \%$  randomisiert mutagenisiert.  $27,4 \% \pm 5,7 \%$  der Synthesefehlerrate wurden auf die Kontamination der Mutagenesebanken mit Template-DNA zurückgeführt. Da dies ein häufiges Problem bei der Erstellung von Mutagenesebanken ist, wurden vor der Mutagenese die zu analysierenden Loci deletiert und mittels spezieller Mutagenese-Primer nachträglich inseriert. Dadurch war stets eine Unterscheidung zwischen Template-DNA (Locus deletiert) und Mutagenese-Produkt (Locus inseriert) möglich. Dennoch reichte die Syntheseeffizienz aus, um eine 100 %ige Motivabdeckung zu erreichen. Der Anteil an Plasmiden in den DL- und PL-Mutagenesebanken mit unvollständigen Mutationskassetten lag bei durchschnittlich 0,32 % und bestätigte die geringe Fehlerrate der Methode.

Neben der hier angewandten PCR-Technik können zur Herstellung von Mutagenesebanken weitere Syntheseverfahren eingesetzt werden (Wang et al. 2004; Singh & Cooper 2006; Kuroyanagi et al. 2010; Ke et al. 2011). Beispielsweise kann über inserierte Restriktionsschnittstellen in das Analyse-Exon eines Minigens ein Austausch des SRMs durch einen randomisierten Oligonukleotid-Pool erfolgen (Singh & Cooper 2006). Hierbei besteht die Möglichkeit einer unvollständigen Restriktion und somit einer Template-Kontamination der Mutagenesebank. Des Weiteren kann das Einfügen von Restriktionsschnittstellen einen Einfluss auf das Spleißgeschehen ausüben, der durch die in dieser Arbeit angewendete Synthesestrategie vermieden wurde.

Zur Identifizierung von SRMs in Abhängigkeit vom Zelltyp wurden drei eukaryotische Zelllinien mit den Mutagenesebanken transient transfiziert. Dabei wurde eine vollständige Motivabdeckung sowohl in HeLa, HuH7 als auch in SH-SY5Y erzielt, bei der jedes Motiv mindestens einmal erfolgreich transfiziert werden konnte. Insgesamt konnten  $97,5 \% \pm 1,5 \%$  der Sequenzen der Mutagenesebanken und der Transkript-Pools der Zelllinien (RT-PCR-Produkte) einer Referenzsequenz zugewiesen werden. Mit einer mindestens 23x Motivabdeckung, wobei eine minimal 5x Abdeckung bereits ausreichend ist, wurden die Mutagenesebanken und Transkript-Pools in einer ausreichenden Tiefe sequenziert.

Die Identifizierung von Enhancer- und Silencer-Sequenzen erfolgte im Anschluss durch die Bestimmung des Verhältnisses der Read-Anzahlen (Anzahlen der sequenzierten Fragmente) von

Transkript und korrespondierendem Plasmid. Zur Validierung der identifizierten SRMs wurde der quantitativ evaluierte Einfluss einiger, zufällig ausgewählter Motive in unabhängigen Zellkulturexperimenten reproduziert ( $R^2 \geq 0,9$ ). Dies bestätigte die Gültigkeit der identifizierten SRMs und zeigt, dass die im Rahmen dieser Dissertation etablierte Methode zur Analyse weiterer spleißregulatorischer Loci verwendet werden kann.

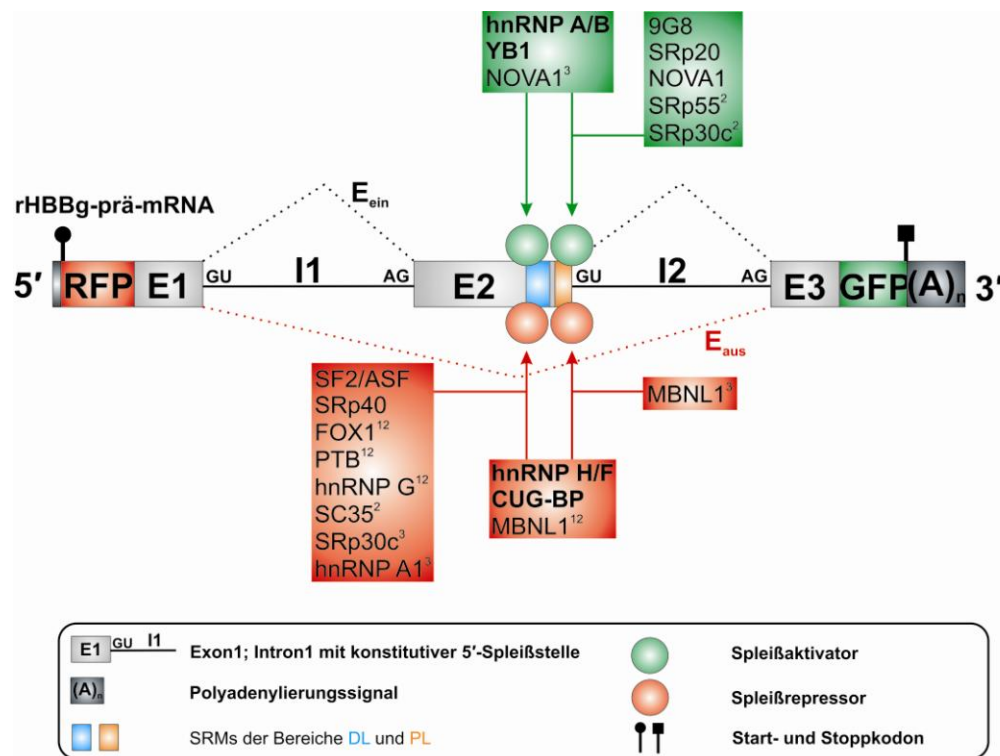
## 6.2 Relevanz der identifizierten SRMs

Da ausschließlich konstitutiv gespleißte Transkripte sequenziert wurden, fiel auf, dass mehr als 50 % der SRMs in der neuronalen Zelllinie SH-SY5Y kein Transkript erzeugten. Bei den HeLa-Zellen, einer epithelialen Zelllinie, waren es hingegen nur ca. 14 % und in der Leberzelllinie HuH7 etwa 30 %. Demnach mussten die nicht detektierten SRMs alternative Spleißreaktionen vermitteln. Vorangegangene Untersuchungen belegen die erhöhte alternative Spleißaktivität der SH-SY5Y-Zellen und postulieren, dass neuronale Zellen mehr als doppelt so viele Exons exprimieren wie Leber- bzw. Epithelzellen (De La Grange et al. 2010).

Um die Bedeutung der SRMs hinsichtlich ihrer Enhancer- und Silencer-Effekte zu bestimmen, wurden diese mithilfe von bereits mathematisch beschriebenen, spleißregulatorischen Faktoren annotiert. Hierfür wurden den SRMs funktionelle Eigenschaften wie GC-Gehalt, Purin-Gehalt, Sekundärstruktur oder Bindepartner zugeordnet. Dabei zeigte sich, dass Silencer mit einem erhöhten Anteil an Sekundärstrukturen und einem erhöhten GC-Gehalt assoziiert sind. Diese beiden Resultate bedingen einander, da ein hoher GC-Gehalt durch die Ausbildung von drei Wasserstoffbrückenbindungen mit einer geringen Entropie und demzufolge mit einem hohen Sekundärstrukturanteil einhergeht. Im Gegensatz dazu weisen Enhancer-Sequenzen einen erhöhten Anteil einzelsträngiger Regionen auf. Die Entropie nahm mit abnehmender Distanz zum Spleißdonor aufgrund der Doppelstranganteile der U1-snRNP-Bindestelle ab, was mit den Ergebnissen anderer Studien übereinstimmte (Krummel et al. 2009). Der Einfluss der Sekundärstruktur der identifizierten Enhancer- und Silencer-Sequenzen (SRMs) auf die Spleißaktivität war in allen drei getesteten Zelllinien konsistent.

Die Bindepartner der identifizierten SRMs wurden mithilfe des Vorhersage-Programms *SFmap* berechnet (Akerman et al. 2009; Paz et al. 2010). Dabei wurde die größtmögliche Sequenzhomologie zwischen den identifizierten SRMs und bereits charakterisierten RNA-Motiven, deren Bindepartner bekannt sind und die in regelmäßigen Abständen aktualisiert werden, ermittelt. Die Gene für die insgesamt 19 charakterisierten RNA-Bindeproteine des Vorhersageprogramms werden in den analysierten Zelllinien exprimiert, wobei Unterschiede im Expressionsniveau existieren (De La Grange et al. 2010). So werden in den SH-SY5Y-Zellen, im Gegensatz zu den HeLa- und HuH7-Zellen, die Spleißfaktoren SC35, SRp20, SRp40, SRp55, SRp30c, hnRNP A1 und NOVA1 stärker exprimiert (Licatalosi & Darnell 2006; Ule & Darnell 2006; De La Grange et

al. 2010). Das komplexe Zusammenspiel der Spleißfaktoren mit den analysierten Loci ist in Abbildung 30 schematisch dargestellt.



**Abb. 30: Schematische Darstellung der Interaktionen von spleißregulatorischen Motiven und Bindepartnern zur Erkennung einer 5'-Spleißstelle:** In Abhängigkeit der Zelllinien HeLa (1), HuH7 (2), SH-SY5Y (3) sind die Interaktionen von spleißregulatorischen Motiven (SRMs) und Bindepartnern in den analysierten distalen (DL) und proximalen (PL) Bereichen des *HBB*-Introns 2 dargestellt. Die Interaktion von Enhancer-Sequenzen und Spleißfaktoren (grün) führt zu einer verstärkten 5'-Spleißstellenerkennung (GU) und demzufolge zu einem Exon-Einschluss (E<sub>ein</sub>). Spleißfaktoren, die an Silencer-Sequenzen gebunden sind (rot), hemmen das Spleißstellensignal und führen zu einem Exon-Ausschluss (E<sub>aus</sub>).

Am auffälligsten ist, dass Spleißaktivatoren, d.h. Spleißfaktoren die einen Enhancer-Effekt vermitteln, überwiegend proximal zum Spleißdonor binden, während Silencer-Effekte in der Regel distal vermittelt werden.

Unabhängig von Locus und Zelltyp wurden die Spleißfaktoren hnRNP A/B und YB1 als Spleißaktivatoren und hnRNP H/F und CUG-BP als Spleißrepressoren identifiziert. Diese Spleißfaktoren stellen in Verbindung mit dem RNA-Sekundärstrukturanteil die Basis einer 5'-Spleißstellenerkennung dar. Dabei weisen die Spleißfaktoren YB1 und hnRNP H/F die stärksten Enhancer- und Silencer-Effekte auf, wobei die Effekte im distalen Bereich in SH-SY5Y-Zellen besonders ausgeprägt sind. In vorangegangenen Studien konnte bereits die spleißaktivierende Funktion von YB1 durch Interaktion mit AC-reichen ESEs nachgewiesen werden (Stickeler et al. 2001; Wei et al. 2012). Bis auf den Spleißfaktor hnRNP A/B vermitteln ausschließlich nicht-hnRNP-Proteine einen Enhancer-Effekt und führen zu einer E<sub>ein</sub>-Reaktion durch Rekrutierung des U1-snRNPs an den nativen Spleißdonor. Eines der interessantesten Ergebnisse ist, dass verhältnismäßig viele SR-Proteine mit einer E<sub>aus</sub>-Reaktion in Verbindung gebracht werden konnten, die üblicherweise mit Enhancer-Effekten assoziiert werden. Detail-

lierte Studien zeigten bereits, dass Hypo- und Hyperphosphorylierungen von SR-Proteinen und hnRNPs antagonistische Effekte auslösen können (Huang & Steitz 2005; Stamm 2008). Beispielsweise phosphoryliert die Glykogensynthase-Kinase-3 das bereits phosphorylierte SR-Protein SC35, welches daraufhin eine  $E_{\text{aus}}$ -Reaktion des Exons 10 des Tau-Proteins vermittelt und zur Pathogenese der Parkinson-Krankheit beiträgt (Avila et al. 2004). Der Phosphorylierungsprozess, dessen spleißregulatorische Funktion bisher unzureichend erforscht ist, wird in Abhängigkeit vom Zelltyp reguliert (Huttlin et al. 2010).

Des Weiteren sind die Auswirkungen des zelltypspezifischen Expressionsniveaus der Spleißfaktoren sichtbar. Aus diesem Grund sind die Spleißfaktoren NOVA1 und hnRNP A1 in der Zelllinie SH-SY5Y stärker in das Spleißgeschehen involviert als in den HeLa- und HuH7-Zellen, was auch andere Arbeiten zeigen (Ule et al. 2006; De La Grange et al. 2010). Die spleißregulatorische Wirkungsweise von hnRNP A1 wurde von Chiou et al. in diesem Jahr aufgeklärt. Eine hnRNP-L-abhängige Bindung von hnRNP A1 in 5'-Spleißstellennähe vermittelt eine Rekrutierung des U1-snRNPs stromaufwärts des Spleißdonors und führt zum Funktionsverlust des U6-snRNPs während der Spleißosom-Assemblierung und damit letztendlich zu einer  $E_{\text{aus}}$ -Reaktion (Chiou et al. 2013).

Eine Gruppierung der annotierten Spleißfaktoren, auf der Grundlage der Art oder Anzahl ihrer funktionellen Domänen, ist nicht möglich. Lediglich die als Spleißaktivatoren identifizierten SR-Proteine 9G8 und SRp20 sind evolutionär konserviert (Shepard & Hertel 2009; Busch & Hertel 2011).

### 6.3 Relevanz weiterer spleißregulatorischer Faktoren

Da mit den Assoziationsstudien maximal 32 % des quantitativ evaluierten Einflusses der identifizierten SRMs auf die Spleißaktivität einer 5'-Spleißstellenerkennung aufgeklärt werden konnte, müssen weitere spleißregulatorische Faktoren berücksichtigt werden.

Eventuell üben alternative Spleißdonoren, die bis zu 20 nt von einer nativen 5'-Spleißstelle entfernt lokalisiert sind und auch als Tandem-5'-Spleißstellen bezeichnet werden, einen zusätzlichen Effekt aus. Dabei sind über ein Drittel der annotierten alternativen Donoren nur 2-10 nt von einem konstitutiven Spleißdonor entfernt (Zavolan et al. 2003; Mardis 2008; Hiller & Platzer 2008). Die abundanteste Tandem-5'-Spleißstelle, mit 8550 Vertretern von 165295 annotierten Tandemdonoren, weist die Nukleotidabfolge GYNGYN auf und vermittelt Transkriptisofomne, die sich in einer Länge von 3 nt unterscheiden (Hiller, Huse, et al. 2006; Hiller, Nikolajewa, et al. 2007; Hiller et al. 2008). Da die Tandemspleißstellen stark degeneriert sind und nicht exakt von Silencer-Sequenzen unterschieden werden können, wurde der Effekt alternativer Donoren nicht quantifiziert. Dennoch zeigt die in dieser Arbeit durchgeführte Analyse der U1-snRNP-Bindestelle, dass in Enhancern die Positionen -3 bis -1 kein GTC, GTG und GTT in allen getesteten Zelllinien aufweisen. Demzufolge existiert eine Tandem-5'-Spleißstelle mit der Konsensus-

sequenz GTBGTG. Lediglich die Zelllinie HeLa wies zusätzlich einen Tandemdonor GTAGTG auf.

Neben den klassischen Enhancer- und Silencer-Effekten üben die Chromatinstruktur und Histonmodifikationen ebenfalls einen spleißregulatorischen Einfluss aus, indem sie die korrekte Assemblierung des katalytisch aktiven B-Komplexes des Spleißosoms beeinflussen (Luco et al. 2011). Auch transient transfizierte DNA wird nach dem Eintritt in den Zellkern mit zellulären Histonen in chromatinähnliche Strukturen überführt (Déry et al. 1985; Luo et al. 1998). Dies ist ein notwendiger Prozess, der zum einen die DNA vor dem zellulären Abbau schützt und zum anderen die physiologische Struktur der Matrize für die RNA-Polymerase II darstellt. Verschiedene Studien zeigten, dass Exons mit einer größeren Nukleosomendichte und individuellen Histonmodifikationen assoziiert sind (Andersson et al. 2009; Tilgner et al. 2009; Luco et al. 2010). Daher ist es möglich, dass einige der SRMs als Enhancer agierten, indem sie die Nukleosom-Assemblierung beschleunigten. Zur Einschätzung des Chromatineffektes könnte nach Kaplan et al. 2009 der *Nukleosom-Belegungs-Score* für jedes SRM berechnet werden (Kaplan et al. 2009).

Einen zusätzlichen spleißregulatorischen Effekt könnte die Dichte der transfizierten Zellen ausgeübt haben. Der Einfluss der Zelldichte wurde bereits an Tandem-3'-Spleißstellen nachgewiesen und besitzt eine koregulatorische Konsistenz zwischen Mensch und Maus (Szafranski et al. 2013, siehe Veröffentlichungen, S. 77 dieser Arbeit). Zur Quantifizierung des Effektes müssten rHBBg-Mutationsstudien in Zellen mit unterschiedlicher Zelldichte durchgeführt und die Transkriptisoverhältnisse bestimmt werden. Da in dieser Doktorarbeit Zellen mit gleichen Konfluenzstadien transfiziert wurden, kann der spleißregulatorische Einfluss der Zelldichte nur marginal sein.

Der vermutlich größte nicht quantifizierbare Effekt besteht in der kombinatorischen Anordnung der DL- und PL-Analysen. Würde ein SRM positiv oder negativ synergistisch mit anderen Motiven im DL bzw. PL-Reporterbereich interagieren, wäre der beobachtete Einfluss auf die 5'-Spleißstellenerkennung höher bzw. niedriger gewesen als angenommen. Zusätzlich hätten Spleißfaktoren, die in den analysierten Loci binden, mit nicht mutierten Regionen, die im Exon und im Intron lokalisiert sind, interagieren können. Diese Synergieeffekte würden zu einer hohen Varianz und Fehlklassifikation führen. Die Anfertigung von zusätzlichen randomisierten Mutagenesebanken verschiedener Loci kann zur Identifizierung von Partnersequenzen der putativen synergistischen Paare des analysierten Kontextes beitragen.

#### 6.4 Gezielter Abbau der Primärtranskripte

Als ein nicht-spleißregulatorischer Effekt konnte der NMD-Mechanismus identifiziert werden. In Säugetieren erfolgt die Erkennung eines PTCs über einen Exon-Junction-Komplex (engl. *exon junction complex*, EJC), dessen Bindung durch das Spleißosom sequenzunspezifisch 20-24 nt

stromaufwärts an jeder Exon-Exon-Grenze (engl. *exon junction*) vermittelt wird (Le Hir et al. 2000; Tange et al. 2005). Terminiert das Ribosom während der ersten Translationsrunde (engl. *pioneer round of translation*) an einem PTC, das sich mehr als 50-55 nt stromaufwärts einer Exon-Exon-Grenze befindet, erfolgt die Bildung des NMD-Komplexes und der enzymatische mRNA-Abbau (Thermann et al. 1998; Schoenberg & Maquat 2012). Um einer NMD vermittelten mRNA-Degradation vorzubeugen, wurde der rHBBg-Reporter unter Berücksichtigung des beschriebenen EJC-Modells konstruiert. Dennoch wurde ein NMD-Effekt identifiziert, der in der HeLa-Zelllinie besonders ausgeprägt war und in der distalen Region eine stärkere Wirkung aufwies als im proximalen Reporterbereich. Zu diesem Ergebnis kamen bereits auch andere Studien (Carter et al. 1996; Wang et al. 2002; Bühler et al. 2004; Linde et al. 2007; Weischenfeldt et al. 2012).

Bislang konnte experimentell gezeigt werden, dass mit zunehmendem Abstand zwischen Stoppkodon und Polyadenylierungssignal die mRNA-Konzentration sinkt und daher der NMD-Effekt zunimmt (Bühler et al. 2006). Infolge dessen existiert neben dem EJC-Modell ein weiterer NMD-Mechanismus. Anhand der Organismen *Saccharomyces cerevisiae* und *Drosophila melanogaster*, die einen weiteren NMD-Mechanismus aufweisen, wurde das sogenannte „vereinheitlichte 3'-UTR-Modell“ abgeleitet (Brognia & Wen 2009). Die Theorie besagt, dass mit zunehmendem Abstand des nativen Stoppkodons zum Polyadenylierungssignal eine Kommunikation des Poly-A-Schwanz-Bindeproteins (PABPC1) und des NMD-Faktors UPF1 zustande kommt. Aus diesem Grund findet keine Interaktion zwischen PABPC1 und dem Translationsterminationsfaktor eRF3 statt, wodurch die eRF3-Ribosom-Assoziation gestört ist, keine kanonische Translationstermination vermittelt wird und der NMD-Mechanismus über ein aktives UPF1 initiiert wird (Ivanov et al. 2008). Da mit dem etablierten Reportersystem auch NMD-Effekte durch fehlende Fluoreszenz gemessen werden können, könnten zusätzliche Mutationsstudien mit PTCs, deren Abstand zum Polyadenylierungssignal variiert, analysiert werden. Dies würde die Bestimmung des Einflusses des Abstandes zwischen PTC und Polyadenylierungssignal auf einen gezielten mRNA-Abbau ermöglichen.

## 6.5 Spleißrelevante Mutationen im humanen Genom

Bisherige bioinformatische Arbeiten, die durchgeführt wurden, um spleißrelevante Mutationen im humanen Genom besser bewerten zu können, erzielten keine zufriedenstellenden Ergebnisse (ElSharawy et al. 2009). Hier könnten die im Rahmen dieser Arbeit identifizierten 1812 Enhancer- und 2803 Silencer-Sequenzen eine verbesserte Interpretation von spleißrelevanten Mutationen gewährleisten.

Die abundantesten genetischen Variationen des humanen Genoms stellen die SNPs dar. Im Jahr 2010 wurden für 21.429 SNPs der damals 15 Mio. bekannten SNPs Krankheitszusammenhänge mittels genetischer Assoziationsstudien vorhergesagt (Durbin et al. 2010; Chen et al.



2010). Aufgrund der technischen Entwicklung in den vergangenen drei Jahren ist die Anzahl der annotierten SNPs auf über 40 Mio. angewachsen und deren Bedeutung als genetischer Marker sowie Krankheitsauslöser gestiegen (NCBI dbSNP build 138 for human, Apr. 25, 2013). Dabei haben SNPs, die das Spleißgeschehen beeinflussen, physiologische Effekte auf nahezu alle humanen Erkrankungen (Sauna & Kimchi-Sarfaty 2011). Bereits über 30 % der monogenetischen Krankheiten können auf SNP-assoziiertes Fehlspleißen zurückgeführt werden (Baralle & Baralle 2005). Aber auch seltenere Mutationen (engl. *rare variants*) in direkter Nachbarschaft zur Exon-Intron-Grenze zeigten Krankheitsrelevanzen (Rosenow et al. 2002).

Im Zuge der Validierung der in dieser Arbeit identifizierten SRMs konnte beobachtet werden, dass schon die Mutation einer einzigen Position in SH-SY5Y-Zellen zu bis zu 24 % mehr alternativ gespleißten Transkripten führt. Somit kann ein Abgleich der SRMs mit annotierten Mutationen zur Identifizierung spleißregulatorischer Bereiche, zur besseren Einschätzung von 5'-Spleißstellenstärken und zur Identifizierung von funktionellen Pseudo-Exons und latenten 5'-Spleißstellen im humanen Genom eingesetzt werden. Auf der Grundlage dieser Ergebnisse könnten krankheitsrelevante SNPs bzw. Mutationen mit genetischen Assoziationsstudien identifiziert werden.

## 6.6 Ausblick

Die Ergebnisse dieser Arbeit haben gezeigt, dass mithilfe der etablierten Methode der fluoreszenzbasierten Unterscheidung zwischen konstitutivem und alternativem Spleißen SRMs schnell und kostengünstig quantitativ evaluiert werden können. Durch die Anwendung dieser Methode könnte in zukünftigen Studien der Einfluss verschiedener Genkontexte und von Mutationen in regulatorischen Bereichen auf das Spleißen untersucht werden. Hierfür wurden schon acht weitere Spleißreporter-Plasmide entsprechend dem rHBBg-System kloniert. Durch einen Vergleich bereits identifizierter und neu identifizierter SRMs, können Unterschiede in der Spleißaktivität quantifiziert und die Bedeutung des jeweiligen Genkontextes bestimmt werden.

Um die biologische Relevanz der ermittelten Bindepartner der in dieser Arbeit identifizierten SRMs zu überprüfen, sollen Protein-RNA-Interaktionen experimentell bestimmt werden. Denkbar wären UV-Crosslinking- und Immunopräzipitationsanalysen (CLIP), aber auch siRNA-Experimente. CLIP ist eine biochemische Methode, die auf der Quervernetzung von RNA-Proteinkomplexen basiert. Werden identifizierte SRMs (RNA-Sequenzbereiche) mit definierten Spleißfaktoren inkubiert und mit dieser Methode analysiert, können so spezifische Bindepartner für die einzelnen SRMs identifiziert werden. Alternativ können einzelne Spleißfaktoren in den Zellen mit einem siRNA-vermittelten Gen-Knockdown herunter reguliert werden. Durch die anschließende fluoreszenzbasierte Quantifizierung der Transkriptisoformen, kann indirekt der Einfluss einzelner Spleißfaktoren auf bestimmte SRMs und die Stärke der Interaktion in den Zellen bestimmt werden.

---

Weiterhin ist die Integration neu identifizierter SRMs in entsprechende Datenbanken für die *in silico* Beurteilung der Spleißrelevanz von Mutationen und für die potentielle Aktivierung von latenten 5'-Spleißstellen denkbar. Somit wären mithilfe der identifizierten SRMs Rückschlüsse auf spleißregulatorisch bedeutsame und krankheitsassoziierte Veränderungen im humanen Genom möglich.

Letztendlich bietet die etablierte Methode für die quantitative Evaluierung von SRMs und deren Bindepartnern einen neuen experimentellen Ansatz, um die Krankheitsrelevanz von Spleißereignissen näher zu untersuchen.

---

## Literaturverzeichnis

- Adams, M.D., 2000. The Genome Sequence of *Drosophila melanogaster*. *Science*, 287: 2185–2195.
- Akbari, A. et al., 2008. Improved DNA fragment length estimation in capillary electrophoresis. *Electrophoresis*, 29: 1273–1285.
- Akerman, M. et al., 2009. A computational approach for genome-wide mapping of splicing factor binding sites. *Genome Biology*, 10: R30.
- Alekseyenko, A. V, Kim, N. & Lee, C.J., 2007. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA*, 13: 661–670.
- Anderson, N.G., Matheson, A. & Anderson, N.L., 2001. Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics*, 1: 3–12.
- Andersson, R. et al., 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Research*, 19: 1732–1741.
- Ast, G., 2004. How did alternative splicing evolve? *Nature Reviews Genetics*, 5: 773–782.
- Avila, J. et al., 2004. Assembly in vitro of tau protein and its implications in Alzheimer's disease. *Current Alzheimer Research*, 1: 97–101.
- Bach, M., Winkelmann, G. & Lührmann, R., 1989. 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proceedings of the National Academy of Sciences of the United States of America*, 86: 6038–6042.
- Baralle, D. & Baralle, M., 2005. Splicing in action: assessing disease causing sequence changes. *Journal of Medical Genetics*, 42: 737–748.
- Barash, Y. et al., 2010. Deciphering the splicing code. *Nature*, 465: 53–59.
- Berget, S.M., 1995. Exon recognition in vertebrate splicing. *The Journal of Biological Chemistry*, 270: 2411–2414.
- Black, D.L., 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, 72: 291–336.
- Brogna, S. & Wen, J., 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology*, 16: 107–113.
- Bühler, M. et al., 2006. EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nature Structural & Molecular Biology*, 13: 462–464.
- Bühler, M., Paillusson, A. & Mühlemann, O., 2004. Efficient downregulation of immunoglobulin mu mRNA with premature translation-termination codons requires the 5'-half of the VDJ exon. *Nucleic Acids Research*, 32: 3304–3315.
- Buratti, E. & Baralle, F.E., 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. *Molecular and Cellular Biology*, 24: 10505–10514.
- Burset, M., Seledtsov, I.A. & Solovyev, V. V, 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research*, 28: 4364–4375.

- Busch, A. & Hertel, K.J., 2011. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews RNA*, 3: 1–12.
- Cartegni, L. et al., 2003. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Research*, 31: 3568–3571.
- Cartegni, L., Chew, S.L. & Krainer, A.R., 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics*, 3: 285–298.
- Cartegni, L. & Krainer, A.R., 2002. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature Genetics*, 30: 377–384.
- Carter, M.S., Li, S. & Wilkinson, M.F., 1996. A splicing-dependent regulatory mechanism that detects translation signals. *The European Molecular Biology Organization Journal*, 15: 5965–5975.
- Chen, R. et al., 2010. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE*, 5: e13574.
- Chiara, M.D. et al., 1997. Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *The European Molecular Biology Organization Journal*, 16: 4746–4759.
- Chiou, N.-T., Shankarling, G. & Lynch, K.W., 2013. hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly. *Molecular Cell*, 49: 972–982.
- Cohen, J., 1992. Quantitative methods in psychology: a power primer. *Psychological Bulletin*, 112: 155–159.
- Collins, C.A. & Guthrie, C., 2001. Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. *RNA*, 7: 1845–1854.
- De Conti, L., Baralle, M. & Buratti, E., 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews RNA*, 4: 49–60.
- Cooke, C. & Alwine, J.C., 2002. Characterization of specific protein-RNA complexes associated with the coupling of polyadenylation and last-intron removal. *Molecular and Cellular Biology*, 22: 4579–4586.
- Cramer, P. et al., 1997. Functional association between promoter structure and transcript alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America*, 94: 11456–11460.
- Cunningham, T.P., Hagan, J.P. & Grabowski, P.J., 1995. Reconstitution of exon-bridging activity with purified U2AF and U1 snRNP components. *Nucleic Acids Symposium Series*, 218–219.
- Déry, C. V et al., 1985. The structure of adenovirus chromatin in infected cells. *The Journal of General Virology*, 66: 2671–2684.
- Deutsch, M. & Long, M., 1999. Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Research*, 27: 3219–3228.
- Durbin, R.M. et al., 2010. A map of human genome variation from population-scale sequencing. *Nature*, 467: 1061–1073.

- EISharawy, A. et al., 2009. Systematic evaluation of the effect of common SNPs on pre-mRNA splicing. *Human Mutation*, 30: 625–632.
- Fairbrother, W.G. et al., 2002. Predictive identification of exonic splicing enhancers in human genes. *Science*, 297: 1007–1013.
- Fairbrother, W.G. & Chasin, L.A., 2000. Human genomic sequences that inhibit splicing. *Molecular and Cellular Biology*, 20: 6816–6825.
- Faustino, N.A. & Cooper, T.A., 2003. Pre-mRNA splicing and human disease. *Genes & Development*, 17: 419–437.
- Fedorov, A. et al., 2001. Comparison of intron-containing and intron-lacking human genes elucidates putative exonic splicing enhancers. *Nucleic Acids Research*, 29: 1464–1469.
- Gao, K. et al., 2008. Human branch point consensus sequence is yUnAy. *Nucleic Acids Research*, 36: 2257–2267.
- Goren, A. et al., 2006. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Molecular Cell*, 22: 769–781.
- Gouloudis, C., 2004. Expression , Charakterisierung und Kristallisation der spleißosomalen Proteine, *Georg-August-Universität zu Göttingen*.
- Graveley, B.R., 2000. Sorting out the complexity of SR protein functions. *RNA*, 6: 1197–1211.
- Hattori, M., 2005. Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso Protein Nucleic Acid Enzyme*, 50: 162–168.
- Heinzen, E.L. et al., 2007. Nova2 interacts with a cis-acting polymorphism to influence the proportions of drug-responsive splice variants of SCN1A. *American journal of human genetics*, 80: 876–883.
- Hiller, M. et al., 2008. Assessing the fraction of short-distance tandem splice sites under purifying selection. *RNA*, 14: 616–629.
- Hiller, M., Huse, K., et al., 2006. Phylogenetically widespread alternative splicing at unusual GYNGYN donors. *Genome Biology*, 7: R65.
- Hiller, M., Zhang, Z., et al., 2007. Pre-mRNA secondary structures influence exon recognition. *PLoS Genetics*, 3: e204.
- Hiller, M., Nikolajewa, S., et al., 2007. TassDB: a database of alternative tandem splice sites. *Nucleic Acids Research*, 35: D188–D192.
- Hiller, M., Pudimat, R., et al., 2006. Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. *Nucleic Acids Research*, 34: e117.
- Hiller, M. & Platzer, M., 2008. Widespread and subtle: alternative splicing at short-distance tandem sites. *Trends in Genetics*, 24: 246–255.
- Hillgren, P. & Parker, R., 1999. Mechanisms of mRNA surveillance in eukaryotes. *Annual Review of Genetics*, 33: 229–260.
- Hillman, R.T., Green, R.E. & Brenner, S.E., 2004. An unappreciated role for RNA surveillance. *Genome Biology*, 5: R8.

- Le Hir, H. et al., 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *The European Molecular Biology Organization Journal*, 19: 6860–6869.
- Hoffman, B.E. & Grabowski, P.J., 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes & Development*, 6: 2554–2568.
- Huang, Y. et al., 2010. CD-HIT: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26: 680–682.
- Huang, Y. & Steitz, J.A., 2005. SRprises along a messenger's journey. *Molecular Cell*, 17: 613–615.
- Hulett, H.R. et al., 1969. Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. *Science*, 166: 747–749.
- Huttlin, E.L. et al., 2010. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*, 143: 1174–1189.
- Imle, P., 2005. Fluorescence-based fragment size analysis. *Methods In Molecular Biology Clifton Nj*, 311: 139–146.
- Index, R., Development, T.R. & Team, C., 2003. The R Environment for Statistical Computing and Graphics, <http://www.r-project.org/>.
- Ivanov, P. V et al., 2008. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *The European Molecular Biology Organization Journal*, 27: 736–747.
- Joshi, M. et al., 2008. Real-time PCR to determine transgene copy number and to quantitate the biolocalization of adoptively transferred cells from EGFP-transgenic mice. *Biotechniques*, 45: 247–258.
- Jumaa, H. & Nielsen, P.J., 1997. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *The European Molecular Biology Organization Journal*, 16: 5077–5085.
- Kaplan, N. et al., 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458: 362–366.
- Ke, S. et al., 2011. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Research*, 21: 1360–1374.
- Keren, H., Lev-Maor, G. & Ast, G., 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics*, 11: 345–355.
- Kibbe, W. a, 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, 35: W43–46.
- Krämer, A., 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annual Review of Biochemistry*, 65: 367–409.
- Krawczak, M., Reiss, J. & Cooper, D.N., 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Human Genetics*, 90: 41–54.

- Krummel, D.A.P. et al., 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 angstrom resolution. *Nature*, 458: 475–480.
- Kuroyanagi, H. et al., 2010. Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic *Caenorhabditis elegans*. *Nature Protocols*, 5: 1495–1517.
- Kwok, J.B.J. et al., 2005. GSK3B polymorphisms alter transcription and splicing in Parkinson's disease. *Annals of Neurology*, 58: 829–839.
- De La Grange, P. et al., 2010. Splicing factor and exon profiling across human tissues. *Nucleic Acids Research*, 38: 2825–2838.
- Lander, E. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409: 860–921.
- Lee, V.M., Goedert, M. & Trojanowski, J.Q., 2001. Neurodegenerative tauopathies. *Annual Review of Neuroscience*, 24: 1121–1159.
- Levine, A. & Durbin, R., 2001. A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Research*, 29: 4006–4013.
- Levine, M. & Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 334: 192–193.
- Li, B. et al., 2002. Stop codons affect 5' splice site selection by surveillance of splicing. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 5277–5282.
- Licatalosi, D.D. & Darnell, R.B., 2006. Splicing regulation in neurologic disease. *Neuron*, 52: 93–101.
- Lin, M.F. et al., 2007. Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Research*: 1823–1836.
- Linde, L. et al., 2007. The efficiency of nonsense-mediated mRNA decay is an inherent character and varies among different cells. *European journal of human genetics*, 15: 1156–1162.
- Lopez, A.J., 1998. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annual Review of Genetics*, 32: 279–305.
- López-Bigas, N. et al., 2005. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters*, 579: 1900–1903.
- Luco, R.F. et al., 2011. Epigenetics in alternative pre-mRNA splicing. *Cell*, 144: 16–26.
- Luco, R.F. et al., 2010. Regulation of alternative splicing by histone modifications. *Science*, 327: 996–1000.
- Luo, R.X., Postigo, A.A. & Dean, D.C., 1998. Rb interacts with histone deacetylase to repress transcription. *Cell*, 92: 463–473.
- Maniatis, T. & Reed, R., 2002. An extensive network of coupling among gene expression machines. *Nature*, 416: 499–506.

- Maniatis, T. & Tasic, B., 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*, 418: 236–243.
- Mardis, E.R., 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9: 387–402.
- Martinez-Contreras, R. et al., 2007. hnRNP proteins and splicing control. *Advances in Experimental Medicine and Biology*, 623: 123–147.
- McKernan, K.J. et al., 2009. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Research*, 19: 1527–41.
- Miriami, E. et al., 2002. Conservation of an open-reading frame as an element affecting 5' splice site selection. *Journal of Structural Biology*, 140: 116–122.
- Miriami, E. et al., 2004. Regulation of splicing: the importance of being translatable. *RNA*, 10: 1–4.
- Modrek, B. et al., 2001. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Research*, 29: 2850–2859.
- Mullis, K. et al., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, 51: 263–273.
- Neves, G. et al., 2004. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nature Genetics*, 36: 240–246.
- Nevo, Y. et al., 2012. Genome-wide activation of latent donor splice sites in stress and disease. *Nucleic Acids Research*, 40: 10980–10994.
- Pan, Q. et al., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40: 1413–1415.
- Paz, I. et al., 2010. SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Research*, 38: W281–W285.
- Proudfoot, N.J., Furger, A. & Dye, M.J., 2002. Integrating mRNA Processing with Transcription. *Cell*, 108: 501–512.
- Robberson, B.L., Cote, G.J. & Berget, S.M., 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and Cellular Biology*, 10: 84–94.
- Roca, X., Sachidanandam, R. & Krainer, A.R., 2005. Determinants of the inherent strength of human 5' splice sites. *RNA*, 11: 683–698.
- Rosenow, C., Maniar, M. & Trias, J., 2002. Distribution and characterization of regulatory elements in the human genome. *Genome Research*, 9: 1189–1197.
- Rozen, S. & Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers S. Krawetz & S. Misener, eds. *Methods in Molecular Biology*, 132: 365–386.
- Sanger, F. & Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94: 441–448.



- Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74: 5463–5467.
- Sauna, Z.E. & Kimchi-Sarfaty, C., 2011. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics*, 12: 683–691.
- Scaffidi, P. & Misteli, T., 2006. Lamin A-dependent nuclear defects in human aging. *Science*, 312: 1059–1063.
- Schaal, T.D. & Maniatis, T., 1999. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Molecular and Cellular Biology*, 19: 261–273.
- Schellenberg, M.J., Ritchie, D.B. & MacMillan, A.M., 2008. Pre-mRNA splicing: a complex picture in higher definition. *Trends in Biochemical Sciences*, 33: 243–246.
- Schindler, S. et al., 2009. Comparison of methods for quantification of subtle splice variants. *Electrophoresis*, 30: 3674–3681.
- Schoenberg, D.R. & Maquat, L.E., 2012. Regulation of cytoplasmic mRNA decay. *Nature Reviews Genetics*, 13: 246–259.
- Schoft, V.K., Schopoff, S. & Jantsch, M.F., 2007. Regulation of glutamate receptor B pre-mRNA splicing by RNA editing. *Nucleic Acids Research*, 35: 3723–3732.
- Shatkin, A.J. & Manley, J.L., 2000. The ends of the affair: capping and polyadenylation. *Nature Structural Biology*, 7: 838–842.
- Shepard, P.J. & Hertel, K.J., 2009. The SR protein family. *Genome Biology*, 10: 242.
- Singh, G. & Cooper, T.A., 2006. Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing. *Biotechniques*, 41: 177–181.
- Solis, A.S., Shariat, N. & Patton, J.G., 2008. Splicing fidelity, enhancers, and disease. *Frontiers in Bioscience*, 13: 1926–1942.
- Somarelli, J.A. et al., 2013. Fluorescence-based alternative splicing reporters for the study of epithelial plasticity in vivo. *RNA*, 19: 116–127.
- Staley, J.P. & Guthrie, C., 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, 92: 315–326.
- Stamm, S. et al., 2006. ASD: a bioinformatics resource on alternative splicing. *Nucleic Acids Research*, 34: D46–D55.
- Stamm, S., 2008. Regulation of alternative splicing by reversible protein phosphorylation. *The Journal of Biological Chemistry*, 283: 1223–1227.
- Stickeler, E. et al., 2001. The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *The European Molecular Biology Organization Journal*, 20: 3821–3830.
- Sugnet, C.W. et al., 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pacific Symposium On Biocomputing*, 77: 66–77.
- Sun, H. & Chasin, L.A., 2000. Multiple splicing defects in an intronic false exon. *Molecular and Cellular Biology*, 20: 6414–6425.

- Sureau, A. et al., 2001. SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *The European Molecular Biology Organization Journal*, 20: 1785–1796.
- Tange, T.Ø. et al., 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA*, 11: 1869–1883.
- Thanaraj, T.A. & Stamm, S., 2003. Prediction and statistical analysis of alternatively spliced exons. *Progress in Molecular and Subcellular Biology*, 31: 1–31.
- Thermann, R. et al., 1998. Binary specification of nonsense codons by splicing and cytoplasmic translation. *The European Molecular Biology Organization Journal*, 17: 3484–3494.
- Tilgner, H. et al., 2009. Nucleosome positioning as a determinant of exon recognition. *Nature Structural & Molecular Biology*, 16: 996–1001.
- Ule, J. et al., 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature*, 444: 580–586.
- Ule, J. & Darnell, R.B., 2006. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Current Opinion in Neurobiology*, 16: 102–110.
- Valentonyte, R. et al., 2005. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nature Genetics*, 37: 357–364.
- Venter, J.C. et al., 2001. The sequence of the human genome. *Science*, 291: 1304–51.
- Vilardell, J. et al., 2000. The odyssey of a regulated transcript. *RNA*, 6: 1773–1780.
- Walbot, V., 2000. Arabidopsis thaliana genome. A green chapter in the book of life. *Nature*, 408: 794–795.
- Wang, G.-S. & Cooper, T.A., 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics*, 8: 749–761.
- Wang, J. et al., 2002. Boundary-independent polar nonsense-mediated decay. *EMBO Reports*, 3: 274–279.
- Wang, Z. et al., 2004. Systematic identification and analysis of exonic splicing silencers. *Cell*, 119: 831–845.
- Wang, Z. & Burge, C.B., 2008. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*, 14: 802–813.
- Wei, W.-J. et al., 2012. YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *Nucleic Acids Research*, 40: 8622–8636.
- Weischenfeldt, J. et al., 2012. Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biology*, 13: R35.
- Wen, J., Chen, Z. & Cai, X., 2013. A biophysical model for identifying splicing regulatory elements and their interactions. *PloS One*, 8: e54885.
- Will, C.L. & Lührmann, R., 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology*, 13: 290–301.
- Willie, E. & Majewski, J., 2004. Evidence for codon bias selection at the pre-mRNA level in eukaryotes. *Trends in Genetics*, 20: 534–538.

- 
- Wirth, B., Brichta, L. & Hahnen, E., 2006. Spinal muscular atrophy and therapeutic prospects. *Progress in Molecular and Subcellular Biology*, 44: 109–132.
- Wortman, J.R. et al., 2003. Annotation of the Arabidopsis Genome. *Plant Physiology*, 132: 461–468.
- Yeo, G. & Burge, C.B., 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Molecular Cell Biology*, 11: 377–94.
- Zavolan, M. et al., 2003. Impact of Alternative Initiation, Splicing, and Termination on the Diversity of the mRNA Transcripts Encoded by the Mouse Transcriptome. *Genome Research*, 13: 1290–1300.
- Zhang, X.H.-F. et al., 2009. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA*, 15: 367–376.
- Zhang, X.H.-F. & Chasin, L.A., 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes & Development*, 18: 1241–1250.
- Zhang, Z. et al., 2008. SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell*, 133: 585–600.
- Zhong, X.-Y. et al., 2009. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Molecular Cell*, 35: 1–10.

## Anhang

### I. Primer-Liste

Die Bezeichnung der Oligonukleotide setzt sich wie folgt zusammen: Einsatzbereich\_Locus-Sublocus.Orientierung. Demnach stellt beispielsweise der Primer *RT\_HBB-E1.f* einen Forward-Primer (*f*) einer RT-PCR (*RT*) dar, mit einer Hybridisierungsstelle im Exon 1 (*E1*) des *HBB*-Gens (*HBB*). Die Einsatzbereiche und Labelungen der Primer sind wie folgt kodiert: C=Klonierung, OM=Ortsgerichtete Mutagenese, SS=Sanger-Sequenzierung, RM=Randomisierte Mutagenese, IS=Illumina-Tiefensequenzierung, RT=RT-PCR, qRT=qRT-PCR, CE=CE-LIF, P=Phosphorylierung, FAM=6-Carboxyfluorescein.

Die Primer sind entsprechend ihrer Verwendung nach geordnet.

Primer-Name	5'-Oligonukleotidsequenz-3'
<b>rHBBg-Klonierung</b>	
C_HBB-E1-NheI.f	GCTAGCGTGCATCTGACTCCTGAGGAGAAGT
C_HBB-E3-Sall.r	GTCGACGTGATACTTGTGGGCCAGGGCATT
OM_HBBupMluI.f	CAGATCCGCTAGCACCATGAAAGGCACGCGTGTGCATCTGACTCCTGAG
OM_HBBupMluI.r	CTCAGGAGTCAGATGCACACGCGTGCCTTTCATGGTGCTAGCGGATCTG
OM_GFP-DelATG.f	CCGGGATCCACCGTGAGCAAGGGC
OM_GFP-DelATG.r	GCCCTTGCTCACGGTGGATCCCGG
C_RFP-ATG-NheI.f	GCTAGCATGGCCTCCTCCGAGGA
C_RFP-DelTAG-MluI.r	ACGCGTCAGGAACAGGTGGTGGCG
<b>Weitere Reporterklonierungen</b>	
C_ACTB-E1-MluI.f	ACGCGTGATGATGATATCGCCGCGCT
C_ACTB-E3-BamHI.r	GGATCCCAGGAAGGAAGGCTGGAAGA
C_APOC3-E2-MluI.f	ACGCGTCAGCCCCGGGTACTCCTT
C_APOC3-E4-Sall.r	GTCGACGGCAGCCACGGCTGAAGTT
C_EIF3D-E7-MluI.f	ACGCGTAAACCCGAGACTCTTCAGTT
C_EIF3D-E9-Sall.r	GTCGACAAAGTCAGAGTTGTCTCTCTTGTC
C_EPCAM-E2-MluI.f	ACGCGTTGTGTCTGTGAAACTACAAGCT
C_EPCAM-E4-Sall.r	GTCGACCCGCAAACCTTTTACTATCATAAGG
C_E2F4-E5-MluI.f	ACGCGTGATACCCTCTTGCCATCC
C_E2F4-E7-Sall.r	GTCGACACCTGTGGGTCTGCCTT
C_LMNA-E10-MluI.f	ACGCGTGAAGTGGCCATGCGCAAGC
C_LMNA-E12-Sall.r	GTCGACCATGATGCTGCAGTTCTGGG
C_MAT1A-E7-MluI.f	ACGCGTGGGGATGCGGGTGTCACT
C_MAT1A-E9-Sall.r	GTCGACAAATACAAGCTTCTGGGAACCT
C_PGM2L1-E10-MluI.f	ACGCGTGAAACATTACCAGGTTTTTA
C_PGM2L1-E12-XhoI.r	CTCGAGATTTCTTATTAGGCTGGCTACT
<b>Sanger-Sequenzierung</b>	
SS_ACTB-I2.f	CCCGCTACCTCTTCTGGTG

SS_APOC3-I3.f	TCAGCCTCCTCTTTCTCTCAC
SS_APOC3-I3.r	GCAGCCTCTCATTTGGAAAG
SS_EIF3D-I7.f	AGATCTTGCCATTGCACTCC
SS_EIF3D-I8.r	GGTCTACCTGGCAAAGTGCT
SS_EPCAM-I3.f	TGAGTGGGGCTGCCTATACT
SS_E2F4-E6.f	GTGCAGAAGTCCAGGGAATG
SS_HBB-E2.f	AGTCCTTTGGGGATCTGTCCACTC
SS_LMNA-I10.f	TTGCTCCCCTTCTCTCTTCT
SS_MAT1A-EI7.f	GAAGGGCAAAGCTTCTGAAA
SS_PGM2L1-I10.r	CATTTCTTCAGAAGGACCAGGA
SS_Pcmv.f	CAAATGGGCGGTAGGCGTGT
SS_RFP.f	CCAAGGTGTACGTGAAGCAC
SS_GFP.r	CTCCATCTTATTGCCCAGGA
SS_GFP-3'-UTR.r	AGGTTTCAGGGGGAGGTGTGGGA
SS_M13.f	GTTGTAAAACGACGGCCAGT
SS_M13.r	CACAGGAAACAGCTATGACC
<b>Deletionskonstruktion</b>	
OM_HBB-E2-L-19bis-12-Del.f	CTGCACTGTGACAAGCTGCACGTAGAAGTTTCAGGGTGAGTCTATGGGA
OM_HBB-E2-L-19bis-12-Del.r	TCCCATAGACTCACCTGAAGTTCTACGTGCAGCTTGTCACAGTGCAG
OM_HBB-E2-L-14bis-7-Del.f	CAAGCTGCACGTGGATCTTCAGGGTGAGTCTATG
OM_HBB-E2-L-14bis-7-Del.r	CATAGACTCACCTGAAGATCCACGTGCAGCTTG
OM_HBB-E3-Del1nt.f	GGCTGGTGTGGCAATGCCCTGGCC
OM_HBB-E3-Del1nt.r	GGCCAGGGCATTGCCACACCAGCC
OM_HBB-E2-L-7bis-1-Del.f	AGCTGCACGTGGATCCTGAGAAGTGAGTCTATGGGACGCTTGATGT
OM_HBB-E2-L-7bis-1-Del.r	ACATCAAGCGTCCCATAGACTCACTTCTCAGGATCCACGTGCAGCT
OM_MCS-Del2nt.f	CCCACAAGTATCACCGACGGTACCGCGG
OM_MCS-Del2nt.r	CCGCGGTACCGTCGGTGATACTTGTGGG
OM_HBB-E2-L-18bis-4-Del.f	TGACAAGCTGCACGTGAGGGTGAGTCTATGGG
OM_HBB-E2-L-18bis-4-Del.r	CCCATAGACTCACCTCACGTGCAGCTTGTC
OM_HBB-E2-L-19bis-1-Del.f	AGCTGCACTGTGACAAGCTGCACGTGTGAGTCTATGGGACGCTTGATGT
OM_HBB-E2-L-19bis-1-Del.r	ACATCAAGCGTCCCATAGACTCACACGTGCAGCTTGTCACAGTGCAGCT
<b>Randomisierte Mutagenese</b>	
RM-HBB-E2-PL-Random.f	P-TGCACGTGGATCCTGAGAANNNNNNNGTGAGTCTATGGGACGCTTGATGTT
RM-HBB-E2-PL-Random.r	P-GCTTGTCACAGTGCAGCTCACTCAGTGT
RM-HBB-E2-DL-Random.f	P-CACTGTGACAAGCTGCACGTNNNNNNNNAGAACTTCAGGGTGAGTCTATGGGA
RM-HBB-E2-DL-Random.r	P-CAGCTCACTCAGTGTGGCAAAGGTGCCCTT
<b>Illumina-Tiefensequenzierung</b>	
IS_HBB-E2-0N-PCR1.f	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAGCTGCACTGTGACAAGCTG
IS_HBB-E2-1N-PCR1.f	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNAGTAGCTGCACTGTGACAAGCTG

IS_HBB-E2-2N-PCR1.f	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNTACAGCTGCACTGTGACAAGCTG
IS_HBB-E2-3N-PCR1.f	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCGAAGCTGCACTGTGACAAGCTG
IS_HBB-E2-4N-PCR1.f	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGAGCTGCACTGTGACAAGCTG
IS_HBB-E2-5N-PCR1.f	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNACAAGCTGCACTGTGACAAGCTG
IS_HBB-I2-PCR1.r	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGCAAGAATTAAACAAAAGAAAACAA
IS_HBB-GFP-PCR1.r	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTGAACAGCTCGGCGCCCTT
IS_IlluminaAdaptorA-PCR2.f	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
IS_IlluminaAdaptorB-PCR2.r	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
<b>RT-PCR</b>	
RT_HBB-E1.f	GTGCATCTGACTCCTGAGGAGAAGT
RT_HBB-E3.r	GTGATACTTGTGGGCCAGGGCATT
<b>CE-LIF</b>	
CE_RFP.f	FAM-GCCACCACCTGTTCTCTGA
CE_HBB-E3.r	GTGATACTTGTGGGCCAGGGCATT
<b>qRT-PCR</b>	
qRT_C5-I28.f	TAGGGCTGCTTCCACCTAGA
qRT_C5-I28.r	GTCATGCACTCCTTTGCTCA
qRT_RFP.f	TCCAAGGTGTACGTGAAGCA
qRT_RFP.r	TAGATGAAGGAGCCGTCCTG
<b>Weiter Mutagenesen</b>	
OM_HBB-I2-DonorCA.f	GATCCTGAGAACTTCAGGCAGAGTCTATGGGACGCTTG
OM_HBB-I2-DonorCA.r	CAAGCGTCCCATAGACTCTGCCTGAAGTTCTCAGGATC
RM_HBB-E2-DL-AnsatzA.f	CTGCACTGTGACAAGCTGCACGTNNNNNNNNNAGAAGTTCAGGGTGAGTCTATGGGA
RM_HBB-E2-DL-AnsatzA.r	TCCCATAGACTCACCTGAAGTTCTNNNNNNNNNACGTGCAGCTTGTCACAGTGCAG
RM_HBB-E2-DL-AnsatzB.f	P-NNNNAGAAGTTCAGGGTGAGTCTATGGGACGCT
RM_HBB-E2-DL-AnsatzB.r	P-NNNNACGTGCAGCTTGTCACAGTGCAGCTCACT
RM_HBB-E2-DL-AnsatzC.f	siehe RM-HBB-E2-8mer-Random.f
RM_HBB-E2-DL-AnsatzC.r	siehe RM-HBB-E2-8mer-Random.r
RM_HBB-E2- L-19 bis-1-Random.f	P-CACTGTGACAAGCTGCACGTNNNNNNNNNNNNNNNNNGTGAGTCTATGGGACGCTTGATGTT
RM_HBB-E2- L-19 bis-1-Random.r	P-CAGCTCACTCAGTGTGGCAAAGGTGCCCTT

## II. Referenzsequenzen des Sequenzabgleichs

Die Referenzsequenzen sind in 5'-3'-Richtung dargestellt. Großbuchstaben symbolisieren Basen exonischen Ursprungs und Kleinbuchstaben charakterisieren Intron-Regionen.

- **Randomisierte DL-Mutagenesebank**

**nativ\_8xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttcc  
ccttcttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacg  
aatgattgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttctttt  
gtttaattcttgc

**nativ\_7xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttccc  
cttcttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacga  
atgattgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttg  
tttaattcttgc

**nativ\_6xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttcccc  
ttcttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaa  
tgattgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgt  
ttaattcttgc

**nativ\_5xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttccct  
tcttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaat  
gattgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgtt  
taattcttgc

**nativ\_4xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttccctt  
cttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaatg  
attgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgtt  
aattcttgc

**nativ\_3xN**

AGCTGCACTGTGACAAGCTGCACGTNNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttcccttc  
ttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaatga  
ttgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgttta  
attcttgc

**nativ\_2xN**

AGCTGCACTGTGACAAGCTGCACGTNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttcccttct  
tttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaatgat  
tgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgttta  
ttcttgc

**nativ\_1xN**

AGCTGCACTGTGACAAGCTGCACGTNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttcccttctt  
ttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaatgatt  
gcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgttta  
tcttgc

**nativ\_0xN (L<sub>19bis-12</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttcccttcttt  
tctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaatgattg

catcagtgtggaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgtttaatt  
cttgc

**kryptisch1 (L<sub>-7bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAAgtgagtcctatgggacgcttgatgttttc

**kryptisch2 (L<sub>-19bis+140</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTtcgtttttagtttcttttatttgctgttcataacaattggt

**kryptisch3 (L<sub>-19bis+162</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTctgtttcataacaattgttttcttttgtttaattcttgcag

**kryptisch4 (L<sub>-19bis+166</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTtcataacaattgttttcttttgtttaattcttgcagatcg

- **DL-Transkript-Pool**

**nativ\_8xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCC  
CATCACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCT  
GGCCACAAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_7xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCC  
ATCACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTG  
GCCACAAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_6xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCA  
TCACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTG  
CCCACAAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_5xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCAT  
CACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGC  
CCACAAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_4xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATC  
ACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCC  
CACAAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_3xN**

AGCTGCACTGTGACAAGCTGCACGTNNNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCA  
CTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCC  
ACAAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_2xN**

AGCTGCACTGTGACAAGCTGCACGTNNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCAC  
TTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCA  
CAAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_1xN**

AGCTGCACTGTGACAAGCTGCACGTNAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACT  
TTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCA  
AAGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_0xN (L<sub>-19bis-12</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTT  
TGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCA  
AGTATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG



**kryptisch\_9xN (L<sub>10bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTG  
GCAAAGAATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAG  
TATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_8xN (L<sub>11bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGG  
CAAAGAATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGT  
ATCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_7xN (L<sub>12bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGC  
AAAGAATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTA  
TCACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_6xN (L<sub>13bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCA  
AAGAATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTAT  
CACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_5xN (L<sub>14bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAA  
AGAATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATC  
ACGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_4xN (L<sub>15bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAA  
GAATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCA  
CGTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_3xN (L<sub>16bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAG  
AATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCAC  
GTCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_2xN (L<sub>17bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGA  
ATTACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCACG  
TCGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_1xN (L<sub>18bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAA  
TTCACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCACGT  
CGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**kryptisch\_0xN (L<sub>19bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAAT  
TCACCCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCACGTC  
GACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**genomisch**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNNAGAACTTCAGGgtgagtctatgggacgcttgatgttttctttcc  
ccttcttttctatggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacg  
aatgattgcatcagtggtgaagtctcaggatcggttttagtttcttttatttgctgttcataacaattgttttctttt  
gtttaattcttgctttcttttttttcttctcgcgaatttttactattatacttaatgccttaacattgtgtataac  
aaaaggaaatatctctgagatacattaagtaacttaaaaaaaactttacacagctctgcctagtagtactattttg  
gaatatatgtgtgcttattttgcatattcataatctccctactttatttttcttttatttttaattgatacataatcat  
tatacatattttatgggttaaagtgtaatgttttaatatgtgtacacataattgaccaaatacagggttaattttgcattt  
gtaatttttaaaaaatgctttcttcttttaataactttttgtttatcttattttctaatactttccctaatactctttt  
ctttcaggggaataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagtgataattttctgggtt  
aaggcaatagcaatatTtctgcatataaataattttctgcatataaattgttaactgatgtaagaggtttcatattgcta  
atagcagctacaatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagct

agggcccttttgctaatacatgttcataacctcttatcttccctcccacagCTCCTGGGCAACGTGCTGGTCTGTGTGCTG  
GCCCATCACTTTGGCAAAGAATTACCCCCACCAGTGACAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGC  
CCTGGCCCCACAAGTATCACGTGACGGTACCGCGGGCCCGGGATCCACCGTGAGCAAGGG

**Artefakt1 (Primer-Dimer)**

AGCTGCACTGTGACAAGCTGTTCAAGATCGGAAGAGCTCGTATGCCGTCTTCTNNNNNNNNNNNNNNNNNN

**Artefakt2**

AGCTGCACTGTGACAAGCTGAATGTTTCATGGGGCGGCCATCCAGAGGGACGCCGTTGTACTGCTT

**Artefakt3**

AGCTGCACTGTGACAAGCTGCACGTGGGCGCCGAGCTGTTCAAGATCGGAAGAGCTCGTATGCCG

**Artefakt4**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAATCCACCGTGAGCAAGGGCGCCGAGCTGT

**Artefakt5**

AGCTGCACTGTGACAAGCTGAGCAAGATTACAGACCTCAAGCTGGCGGCCAGGTACATCGACTTC

**Artefakt6**

AGCTGCACTGTGACAAGCTGCACGTTCACCGTGAGCAAGGGCGCCGAGCTGTTCAAGATCGGAA

**Artefakt7**

AGCTGCACTGTGACAAGCTGAGGGCGCCGAGCTGTTCAAGATCGGAAGAGCTCGTATGCCGTCTT

**Artefakt8**

AGCTGCACTGTGACAAGCTGCACCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTT

**Artefakt9**

AGCTGCACTGTGACAAGCTGGCCGAGCAGGCCGAGCGCTACGACGACATGGCCACCTGCATGAAG

**Artefakt10**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGGGCGCCGAGCTGTTCAAGATCGGAAGAGCT

**Artefakt11**

AGCTGCACTGTGACAAGCTGCGGTATCCAGGCGGCTCGGGCCTGCTTTGAACACTCTAATTTTTT

**Artefakt12**

AGCTGCACTGTGACAAGCTGCACGTGAGCAAGGGCGCCGAGCTGTTCAAGATCGGAAGAGCTCGT

**Artefakt13**

AGCTGCACTGTGACAAGCTGCACGTGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTG

- Randomisierte PL-Mutagenesebank**

**nativ\_7xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNNNgtgagtctatgggacgcttgatgttttctttcc  
ccttcttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacg  
aatgattgcatcagtggtgaagtctcaggatcgtttagtttcttttatttgctgttcataacaattgttttctttt  
gtttaattcttgc

**nativ\_6xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNNgtgagtctatgggacgcttgatgttttctttccc  
cttcttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacga  
atgattgcatcagtggtgaagtctcaggatcgtttagtttcttttatttgctgttcataacaattgttttcttttg  
tttaattcttgc

**nativ\_5xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNNgtgagtctatgggacgcttgatgttttctttcccc  
ttcttttctatgggttaagttcatgtcataggaaggggataagtaacaggggtacagtttagaatgggaaacagacgaa

tgattgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgt  
ttaattcttgc

**nativ\_4xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNGtgagtctatgggacgcttgatgttttctttccct  
tcttttctatgggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaat  
gattgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgtt  
taattcttgc

**nativ\_3xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNGtgagtctatgggacgcttgatgttttctttccctt  
cttttctatgggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatg  
attgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgtt  
aattcttgc

**nativ\_2xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNGtgagtctatgggacgcttgatgttttctttcccttc  
ttttctatgggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatga  
ttgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgttta  
attcttgc

**nativ\_1xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNGtgagtctatgggacgcttgatgttttctttcccttct  
tttctatgggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatgat  
tgcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgttta  
ttcttgc

**nativ\_0xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAAGtgagtctatgggacgcttgatgttttctttcccttctt  
ttctatgggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatgatt  
gcatcagtggtgaagtctcaggatcgtttttagtttcttttatttgctgttcataacaattgttttcttttgtttaat  
tcttgc

**kryptisch1 (L<sub>-19bis-12</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTAGAACTTCAGGgtgagtctatgggacgcttgatgttttct

**kryptisch2 (L<sub>-7bis+140</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAAtcgtttttagtttcttttatttgctgttc

**Artefakt1 (Primer-Self-Annealing)**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAAGTGGATCCTGAGAAGTGAGTCTATGGGA

**Artefakt2 (Primer-Self-Annealing)**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAAGTGAGCTGCACTGTGACAAGCTGCACGT

**Artefakt3 (Primer-Self-Annealing)**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAAGTGACAAGCTGCACGTGGATCCTGAGAA

- **PL-Transkript-Pool**

**nativ\_7xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCC  
CATCACTTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCT  
GGCCACAAGTATCACGTCGACGGTACCGCGGGCCCGGGATCCACCGTGAGCAAGGG

**nativ\_6xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCC  
ATCACTTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTG  
GCCACAAGTATCACGTCGACGGTACCGCGGGCCCGGGATCCACCGTGAGCAAGGG

**nativ\_5xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCA  
TCACTTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGG  
CCCACAAGTATCACGTGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_4xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCAT  
CACTTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGC  
CCACAAGTATCACGTGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_3xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATC  
ACTTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCC  
CACAAGTATCACGTGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_2xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCA  
CTTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCC  
ACAAGTATCACGTGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_1xN**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCAC  
TTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCA  
CAAGTATCACGTGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**nativ\_0xN (L<sub>7bis-1</sub> deletiert)**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACT  
TTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCA  
AAGTATCACGTGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**genomisch**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAANNNNNNNgtgagtcctatgggacgcttgatgttttctttcc  
ccttcttttctatggttaagttcatgtcataggaaggggataagtaacagggtagagtttagaatgggaaacagacg  
aatgattgcatcagtggtgaagtctcaggatcggttttagtttcttttatttgctgttcataacaattgttttctttt  
gtttaattcttgcttttcttttttttcttctccgcaatttttactattatacttaatgccttaacattgtgtataac  
aaaaggaaatatctctgagatacattaagtaacttaaaaaaaactttacacagtcctagtagtactatttg  
gaatataatgtgtgcttatttgcatattcataatctccctactttattttcttttatttttaattgatacataatcat  
tatacatatttatgggttaaagtgtaatgttttaatatgtgtacacatatattgaccaaatacagggtaattttgcat  
gtaatttttaaaaaatgctttcttcttttaataatactttttgtttatcttattttctaataactttccctaatactctt  
ctttcaggggcaataatgatacaatgtatcatgcctcttttgacaccattctaaagaataacagtgataatttctgggtt  
aaggcaatagcaatatTtctgcatataaaatatttctgcatataaaattgtaactgatgtaagagggtttcatattgcta  
atagcagctacaatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagct  
aggcccttttgcataatcatgttcataaccttattcttctccacagCTCCTGGGCAACGTGCTGGTCTGTGTGCTG  
GCCCATCACTTTGGCAAAGAATTACCCCCACCAAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGC  
CCTGGCCCAAGTATCACGTGACGGTACCGCGGGCCCCGGGATCCACCGTGAGCAAGGG

**Artefakt1 (Primer-Dimer)**

AGCTGCACTGTGACAAGCTGTTCAAGATCGGAAGAGCTCGTATGCCGTCTTCTNNNNNNNNNNNNNNNNNN

**Artefakt2**

AGCTGCACTGTGACAAGCTGAATGTTTCATGGGGCGGCCATCCAGAGGGACGCCGTTGTACTGCTT

**Artefakt3**

AGCTGCACTGTGACAAGCTGCACGTGGGCGCCGAGCTGTTCAAGATCGGAAGAGCTCGTATGCCG

**Artefakt4**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAATCCACCGTGAGCAAGGGCGCCGAGCTGT

**Artefakt5**

AGCTGCACTGTGACAAGCTGAGCAAGATTTCAGACCCTCAAGCTGGCGGCCAGGTACATCGACTTC

**Artefakt6**

AGCTGCACTGTGACAAGCTGCACGTTCCACCGTGAGCAAGGGCGCCGAGCTGTTCAAGATCGGAA

**Artefakt7**

AGCTGCACTGTGACAAGCTGAGGGCGCCGAGCTGTTCAAGATCGGAAGAGCTCGTATGCCGTCTT

**Artefakt8**

AGCTGCACTGTGACAAGCTGCACCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTT

**Artefakt9**

AGCTGCACTGTGACAAGCTGGCCGAGCAGGCCGAGCGCTACGACGACATGGCCACCTGCATGAAG

**Artefakt10**

AGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGGGCGCCGAGCTGTTCAAGATCGGAAGAGCT

**Artefakt11**

AGCTGCACTGTGACAAGCTGCGGTATCCAGGCGGCTCGGGCCTGCTTTGAACACTCTAATTTTTT

**Artefakt12**

AGCTGCACTGTGACAAGCTGCACGTGAGCAAGGGCGCCGAGCTGTTCAAGATCGGAAGAGCTCGT

**Artefakt13**

AGCTGCACTGTGACAAGCTGCACGTGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTG

- **Randomisierte Mutagenesebank des Bereichs L<sub>-19bis-1</sub>**

**nativ\_19xN**

AGCTGCACTGTGACAAGCTGCACGTNNNNNNNNNNNNNNNNNNNNgtgagtctatgggacgcttgatgttttctttcc  
 ccttcttttctatgggttaagttcatgtcataggaaggggataagtaacagggtagagtttagaatgggaaacagacg  
 aatgattgcatcagtggtgaagtctcaggatcgtttagtttcttttatttgctgttcataacaattgttttctttt  
 gtttaattcttgc

**III. Quantifizierte Transkriptisoformen zur Identifizierung spleißregulatorischer Loci**

rHBBg-Plasmid	WA [bp]	HeLa			HuH7			SH-SY5Y		
		Ein	SD	p-Wert (t-Test)	Ein	SD	p-Wert (t-Test)	Ein	SD	p-Wert (t-Test)
WT	0	0,9793	0,0164	-	0,9695	0,0112	-	0,9884	0,0019	-
L <sub>-19bis-12</sub> (Δ8)	7	0,9359	0,0203	0,0447	0,8700	0,0099	0,0003	0,9392	0,0039	3,8493E-05
L <sub>-14bis-7</sub> (Δ8)	4	0,9845	0,0099	0,6620	0,9808	0,0055	0,1932	0,9879	0,0014	0,7375
L <sub>-7bis-1</sub> (Δ7)	5	0,9127	0,0364	0,0444	0,8768	0,0176	0,0015	0,9288	0,0080	0,0002
L <sub>-18bis-4</sub> (Δ15)	9	0,9683	0,0215	0,5178	0,9213	0,0524	0,1949	0,9709	0,0018	0,0003
L <sub>-19bis-1</sub> (Δ19)	17	0,6439	0,0820	0,0023	0,5466	0,0573	0,0002	0,5718	0,0629	0,0003
DoMu	2	0,0084	0,0031	5,7784E-08	0,0073	0,0068	2,2942E-08	0,0051	0,0012	1,8250E-11

#### IV. Quantifizierte Proteinisoformen zur Identifizierung spleißregulatorischer Loci

rHBBg-Plasmid	WA [bp]	HeLa			HuH7			SH-SY5Y		
		Ein	SD	p-Wert (t-Test)	Ein	SD	p-Wert (t-Test)	Ein	SD	p-Wert (t-Test)
WT	0	0,9664	0,0296	-	0,9802	0,0157	-	0,9773	0,0166	-
L-19bis-12 ( $\Delta 8$ )	7	0,8941	0,0258	0,0331	0,8281	0,0299	0,0015	0,8853	0,0491	0,0371
L-14bis-7 ( $\Delta 8$ )	4	0,9522	0,0337	0,6121	0,9680	0,0153	0,3916	0,9517	0,0200	0,1624
L-7bis-1 ( $\Delta 7$ )	5	0,8398	0,0485	0,0182	0,8328	0,0362	0,0029	0,8817	0,0364	0,0144
L-18bis-4 ( $\Delta 15$ )	9	0,9425	0,0258	0,3514	0,9183	0,0191	0,0122	0,9628	0,0271	0,4732
L-19bis-1 ( $\Delta 19$ )	17	0,5299	0,0424	0,0001	0,4928	0,0601	0,0002	0,5405	0,0415	0,0001
DoMu	2	0,0515	0,0357	4,3704E-06	0,0567	0,0416	3,5510E-06	0,0686	0,0408	3,6609E-06

#### V. Vergleich der quantifizierten Transkript- und Proteinisoformen

rHBBg-Plasmid	p-Wert (t-Test)		
	HeLa	HuH7	SH-SY5Y
WT	0,5443	0,3910	0,3135
L-19bis-12 ( $\Delta 8$ )	0,0914	0,0826	0,1308
L-14bis-7 ( $\Delta 8$ )	0,1857	0,2472	0,0351
L-7bis-1 ( $\Delta 7$ )	0,1057	0,1306	0,0942
L-18bis-14 ( $\Delta 15$ )	0,2551	0,9297	0,6318
L-19bis-1 ( $\Delta 19$ )	0,0993	0,3247	0,5121
DoMu	0,1055	0,1121	0,0545

#### VI. Plasmidstandardkurve zur Berechnung der Plasmidkopienzahl

Plasmid	Plasmidmasse [ng]	analysierte Plasmidmasse [ng]	Kopienzahl	log Kopienzahl	C <sub>t</sub> -Wert
rHBBg-Wildtypplasmid	6,96E-09	3	4,31E+08	8,6345	7,563
		0,3	4,31E+07	7,6345	11,06
		0,03	4,31E+06	6,6345	13,91
		0,003	4,31E+05	5,6345	17,23
		0,0003	4,31E+04	4,6345	21,04
		0,00003	4,31E+03	3,6345	24,9
Negativkontrolle (H <sub>2</sub> O)	-	0	0,00E+00	-	>36
C5-pCR2.1-TOPO	4,19E-09	1,81	4,32E+08	8,6355	7,922
		0,181	4,32E+07	7,6355	11,43
		0,0181	4,32E+06	6,6355	15
		0,00181	4,32E+05	5,6355	17,78
		0,000181	4,32E+04	4,6355	22,51
		0,0000181	4,32E+03	3,6355	26,19
Negativkontrolle (H <sub>2</sub> O)	-	0	0,00E+00	-	>36

## VII. Sequenzabgleich und Read-Anzahlbestimmung der Mutagenesebanken

### • Randomisierte DL-Mutagenesebank

Referenzsequenz	Read-Anzahl	
	absolut	relativ [%]
gesamt	14.646.111	100,00
• nativ_8xN	9.630.545	65,75
➤ Filterkriterium bestanden	9.582.954	-
• nativ_7xN	89.816	0,61
• nativ_6xN	19.172	0,13
• nativ_5xN	11.868	0,08
• nativ_4xN	8.760	0,06
• nativ_3xN	9.848	0,07
• nativ_2xN	9.979	0,07
• nativ_1xN	9.830	0,07
• nativ_0xN	4.613.620	31,50
• kryptisch1	2.780	0,02
• kryptisch2	72.704	0,50
• kryptisch3	3.529	0,02
• kryptisch4	6.472	0,04
• nicht aligniert	157.188	1,07

### • Randomisierte PL-Mutagenesebank

Referenzsequenz	Read-Anzahl	
	absolut	relativ [%]
gesamt	14.676.618	100,00
• nativ_7xN	9.737.482	66,35
➤ Filterkriterium bestanden	9.696.157	-
• nativ_6xN	357.297	2,43
• nativ_5xN	40.778	0,28
• nativ_4xN	18.840	0,13
• nativ_3xN	12.337	0,08
• nativ_2xN	11.092	0,08
• nativ_1xN	9.578	0,07
• nativ_0xN	4.313.471	29,39
• kryptisch1	4.273	0,03
• kryptisch2	12.519	0,09
• Artefakt1	42.128	0,29
• Artefakt2	5.863	0,04
• Artefakt3	5.782	0,04
• nicht aligniert	105.178	0,72

### VIII. Motivabdeckung der transfizierten Mutagenesebanken

Grundlagen der Berechnung	Mutagenesebank					
	DL			PL		
rHBBg-Plasmidmasse [ng]	6,95E-09					
Motivmöglichkeiten (4 <sup>n-mer</sup> )	65536			16384		
benötigte Plasmidmenge für 1x Motivabdeckung [ng]	4,56E-04			1,14E-04		
Einbezug der Synthesefehlerrate	DL			PL		
eingesetzte Plasmidmenge pro Transfektion [ng]	2000			2000		
Synthesefehlerrate der Mutagenesebank [%]	34,25			33,65		
• effektiv eingesetzte Plasmidmenge [ng]	1315			1327		
Einbezug der Transfektionseffizienzen	HeLa		HuH7		SH-SY5Y	
	DL	PL	DL	PL	DL	PL
durchschnittliche Transfektionseffizienz [%]	35,48	36,14	13,88	16,98	7,89	9,42
• effektiv transfizierte Plasmidmenge [ng]	467	480	183	225	104	125
➤ Motivabdeckung	1,02E+06	4,21E+06	4,01E+05	1,98E+06	2,28E+05	1,10E+06

### IX. Sequenzabgleich und Read-Anzahlbestimmung der Transkript-Pools

#### • DL-Transkript-Pool

Referenzsequenz	Read-Anzahl											
	HeLa				HuH7				SH-SY5Y			
	Replik 1		Replik 2		Replik 1		Replik 2		Replik 1		Replik 2	
	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]
gesamt	4.702.344	100,00	5.792.449	100,00	5.657.002	100,00	6.981.254	100,00	8.723.912	100,00	9.664.604	100,00
• nativ_8xN	2.753.515	58,56	4.057.869	70,05	1.674.802	29,61	1.586.922	22,73	2.535.201	29,06	2.127.411	22,01
➤ Filterkriterium bestanden	2.716.708	-	4.014.719	-	1.627.014	-	1.546.055	-	2.483.922	-	2.095.238	-
• nativ_7xN	15.611	0,33	22.076	0,38	106.432	1,88	104.908	1,50	38.880	0,45	21.039	0,22
• nativ_6xN	10.187	0,22	13.461	0,23	84.584	1,50	78.153	1,12	26.256	0,30	15.448	0,16
• nativ_5xN	13.291	0,28	20.160	0,35	54.658	0,97	52.728	0,76	22.495	0,26	16.687	0,17
• nativ_4xN	6.246	0,13	10.179	0,18	35.884	0,63	34.542	0,49	20.683	0,24	10.171	0,11
• nativ_3xN	7.956	0,17	10.111	0,17	43.420	0,77	38.937	0,56	13.540	0,16	10.064	0,10
• nativ_2xN	11.354	0,24	18.000	0,31	45.464	0,80	47.957	0,69	23.370	0,27	17.144	0,18
• nativ_1xN	6.998	0,15	9.874	0,17	41.676	0,74	38.974	0,56	16.629	0,19	9.819	0,10
• nativ_0xN	291.016	6,19	387.032	6,68	2.022.699	35,76	2.046.918	29,32	958.019	10,98	702.741	7,27
• kryptisch_9xN	3.428	0,07	6.052	0,10	11.183	0,20	11.461	0,16	7.980	0,09	4.089	0,04
• kryptisch_8xN	5.660	0,12	8.169	0,14	17.510	0,31	18.496	0,26	8.534	0,10	7.142	0,07
• kryptisch_7xN	3.857	0,08	6.139	0,11	6.284	0,11	8.484	0,12	5.577	0,06	3.290	0,03
• kryptisch_6xN	1.405	0,03	1.348	0,02	5.821	0,10	6.712	0,10	3.738	0,04	1.984	0,02
• kryptisch_5xN	8.808	0,19	13.868	0,24	17.492	0,31	17.984	0,26	11.474	0,13	7.806	0,08
• kryptisch_4xN	6.581	0,14	12.498	0,22	23.222	0,41	21.511	0,31	9.774	0,11	8.229	0,09
• kryptisch_3xN	2.089	0,04	2.593	0,04	5.333	0,09	4.748	0,07	3.695	0,04	1.353	0,01
• kryptisch_2xN	3.174	0,07	5.382	0,09	11.128	0,20	8.856	0,13	6.074	0,07	2.608	0,03
• kryptisch_1xN	4.421	0,09	5.909	0,10	6.743	0,12	5.626	0,08	5.005	0,06	5.831	0,06
• kryptisch_0xN	1.161	0,02	2.253	0,04	4.157	0,07	4.222	0,06	1.872	0,02	1.804	0,02



• genomisch	75	0,00	30	0,00	28	0,00	260	0,00	36	0,00	419	0,00
• Artefakt1	1.350.589	28,72	706.922	12,20	1.026.370	18,14	2.460.508	35,24	4.698.186	53,85	6.422.126	66,45
• Artefakt2	1.436	0,03	347	0,01	2.492	0,04	3.602	0,05	19.395	0,22	38.298	0,40
• Artefakt3	4.994	0,11	34.403	0,59	29.005	0,51	19.467	0,28	18.570	0,21	29.012	0,30
• Artefakt4	0	0,00	8	0,00	357	0,01	1	0,00	13.336	0,15	15.680	0,16
• Artefakt5	99	0,00	67	0,00	2	0,00	5	0,00	0	0,00	0	0,00
• Artefakt6	5.829	0,12	66.357	1,15	23.811	0,42	18.240	0,26	9.680	0,11	10.068	0,10
• Artefakt7	1.332	0,03	15.303	0,26	3.598	0,06	11.062	0,16	1.863	0,02	4.451	0,05
• Artefakt8	20.695	0,44	34.996	0,60	17.666	0,31	18.153	0,26	21.829	0,25	17.315	0,18
• Artefakt9	399	0,01	4.368	0,08	1.894	0,03	1.233	0,02	8.785	0,10	717	0,01
• Artefakt10	0	0,00	18	0,00	268	0,00	387	0,01	0	0,00	0	0,00
• Artefakt11	0	0,00	19	0,00	235	0,00	112	0,00	1.368	0,02	166	0,00
• Artefakt12	1.609	0,03	27.113	0,47	7.767	0,14	13.012	0,19	6.057	0,07	3.215	0,03
• Artefakt13	5.905	0,13	10.481	0,18	28.803	0,51	31.714	0,45	11.720	0,13	11.775	0,12
• nicht aligniert	152.624	3,25	279.044	4,82	296.214	5,24	265.359	3,80	194.291	2,23	136.702	1,41

## • PL-Transkript-Pool

Referenzsequenz	Read-Anzahl											
	HeLa				HuH7				SH-SY5Y			
	Replik 1		Replik 2		Replik 1		Replik 2		Replik 1		Replik 2	
	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]	absolut	relativ [%]
gesamt	5.727.090	100,00	6.018.403	100,00	4.937.094	100,00	5.968.318	100,00	10.780.838	100,00	7.778.080	100,00
• nativ_7xN	2.765.773	48,29	3.129.243	51,99	1.543.529	31,26	1.132.894	18,98	1.606.733	14,90	1.682.719	21,63
➤ Filterkriterium bestanden	2.734.282	-	3.091.261	-	1.507.618	-	1.109.487	-	1.581.943	-	1.663.706	-
• nativ_6xN	32.135	0,56	38.909	0,65	75.089	1,52	60.170	1,01	22.622	0,21	25.753	0,33
• nativ_5xN	26.918	0,47	27.709	0,46	212.494	4,30	170.632	2,86	28.203	0,26	19.259	0,25
• nativ_4xN	33.509	0,59	37.843	0,63	86.486	1,75	66.187	1,11	23.624	0,22	24.561	0,32
• nativ_3xN	36.704	0,64	41.407	0,69	63.450	1,29	54.372	0,91	36.225	0,34	25.530	0,33
• nativ_2xN	16.009	0,28	19.096	0,32	42.125	0,85	38.658	0,65	20.898	0,19	12.015	0,15
• nativ_1xN	81.500	1,42	101.986	1,69	69.814	1,41	59.655	1,00	37.645	0,35	49.405	0,64
• nativ_0xN	246.593	4,31	348.002	5,78	1.650.053	33,42	1.289.180	21,60	493.465	4,58	529.637	6,81
• genomisch	910	0,02	1.761	0,03	711	0,01	405	0,01	1.176	0,01	404	0,01
• Artefakt1	2.314.916	40,42	2.023.501	33,62	938.631	19,01	2.834.408	47,49	8.248.098	76,51	5.202.644	66,89
• Artefakt2	1.983	0,03	714	0,01	6.149	0,12	22.995	0,39	27.975	0,26	44.906	0,58
• Artefakt3	3.963	0,07	14.682	0,24	24.020	0,49	12.468	0,21	22.453	0,21	24.000	0,31
• Artefakt4	5.232	0,09	19.275	0,32	4.774	0,10	2	0,00	15.043	0,14	4.426	0,06
• Artefakt5	134	0,00	41	0,00	1	0,00	3	0,00	12.762	0,12	15.508	0,20
• Artefakt6	390	0,01	656	0,01	1.394	0,03	0	0,00	0	0,00	0	0,00
• Artefakt7	911	0,02	10.698	0,18	2.568	0,05	2.798	0,05	1.463	0,01	1.633	0,02
• Artefakt8	4.869	0,09	10.386	0,17	5.725	0,12	5.093	0,09	5.517	0,05	4.786	0,06
• Artefakt9	431	0,01	3.579	0,06	9.935	0,20	9.921	0,17	4.146	0,04	1.052	0,01
• Artefakt10	135	0,00	304	0,01	4.238	0,09	267	0,00	249	0,00	1.532	0,02
• Artefakt11	17	0,00	224	0,00	3.025	0,06	20.949	0,35	486	0,00	5	0,00
• Artefakt12	149	0,00	1.660	0,03	109	0,00	897	0,02	1.132	0,01	746	0,01
• Artefakt13	981	0,02	2.318	0,04	1.750	0,04	2.494	0,04	607	0,01	1.300	0,02
• nicht aligniert	152.928	2,67	184.409	3,06	191.024	3,87	183.870	3,08	170.316	1,58	106.259	1,37

## X. Häufigkeitsverteilungen der Motive in den Transkript-Pools

### • Clusterzentren der DL- und PL-Transkript-Pools

Clusterzentren	log <sub>2</sub> Read-Anzahl pro Motiv					
	HeLa		HuH7		SH-SY5Y	
	Replik 1	Replik 2	Replik 1	Replik 2	Replik 1	Replik 2
DL-Transkript-Pool						
• Cluster 1	0,6508	0,9244	0,2737	0,1947	0,3464	0,2587
• Cluster 2	5,7871	6,4493	5,7110	5,5252	7,8988	8,1489
PL-Transkript-Pool						
• Cluster 1	1,5172	1,5076	0,5867	0,3419	0,4680	0,4077
• Cluster 2	7,2493	7,3717	7,2372	7,8636	9,2091	8,5611

### • Motivverteilung pro Cluster in den DL- und PL-Transkript-Pools

Cluster	Motivanzahl pro Cluster					
	HeLa		HuH7		SH-SY5Y	
	Replik 1	Replik 2	Replik 1	Replik 2	Replik 1	Replik 2
DL-Transkript-Pool						
• Cluster 1	29.603	31.978	40.282	38.128	57.304	59.394
• Cluster 2	35.433	33.058	24.754	26.908	7.732	5.642
PL-Transkript-Pool						
• Cluster 1	3.259	3.115	8.303	12.214	14.016	12.707
• Cluster 2	12.971	13.115	7.927	4.016	2.214	3.523

## XI. Identifizierung des NMD-Effektes

### • HeLa-Zellen

Position im <i>HBB</i> - Intron 2	Replik 1				Replik 2			
	IQR pPTC-	IQR pPTC+	Median- differenz	p-Wert (U-Test)	IQR pPTC-	IQR pPTC+	Median- differenz	p-Wert (U-Test)
-19	4,529	4,702	1,094	0,0547	4,94	5,25	1,31	0,0806
-18	4,518	4,349	2,674	7,2457E-176	4,92	4,59	2,60	6,7166E-182
-17	4,526	4,795	0,164	0,2147	4,94	5,08	0,36	0,1662
-16	4,527	4,872	0,239	0,2767	4,94	5,12	0,84	0,0911
-15	4,539	3,959	2,354	1,1759E-131	4,92	4,37	2,66	1,8774E-190
-14	4,538	4,347	0,703	1	4,94	4,90	1,21	1
-13	4,529	4,454	0,818	1	4,94	4,93	1,16	1
-12	4,363	4,441	2,807	0	4,75	4,74	3,27	0
-7	2,402	3,819	0,019	0,6086	2,37	3,17	0,07	0,8296
-6	2,303	4,977	2,070	4,3287E-112	2,23	4,86	2,13	1,1681E-115
-5	2,443	2,326	0,178	0,9782	2,39	2,18	0,00	0,7267
-4	2,438	2,380	0,302	1	2,39	2,25	0,21	1
-3	2,291	5,279	2,796	6,1173E-139	2,21	5,50	3,67	3,3330E-152

- **HuH7-Zellen**

Position im <i>HBB</i> - Intron 2	Replikat 1				Replikat 2			
	IQR pPTC-	IQR pPTC+	Median- differenz	p-Wert (U-Test)	IQR pPTC-	IQR pPTC+	Median- differenz	p-Wert (U-Test)
-19	4,690	4,272	0,064	0,9947	4,641	4,569	0,065	1
-18	4,677	4,241	0,445	3,7093E-24	4,651	4,367	0,481	1,7813E-17
-17	4,662	4,596	0,259	0,0603	4,634	4,647	0,213	0,1442
-16	4,667	4,692	0,250	0,0843	4,633	4,674	0,249	0,1233
-15	4,672	4,308	0,445	2,1232E-17	4,655	4,240	0,551	6,6546E-22
-14	4,667	4,678	0,030	0,9916	4,642	4,644	0,016	1
-13	4,634	4,837	0,030	0,8641	4,619	4,757	0,033	0,9284
-12	4,641	4,715	0,375	4,7363E-95	4,631	4,698	0,308	6,0485E-77
-7	6,232	6,295	1,216	0,6701	4,253	1,291	0,131	1
-6	6,222	6,100	1,767	2,8836E-13	3,416	1,931	0,148	0,0011
-5	6,200	6,698	1,108	0,0568	3,060	6,097	0,310	0,0601
-4	6,219	6,483	0,798	0,5604	3,464	1,963	0,005	0,9654
-3	6,216	5,839	2,489	8,2151E-26	4,841	1,619	0,168	0,0001

- **SH-SY5Y-Zellen**

Position im <i>HBB</i> - Intron 2	Replikat 1				Replikat 2			
	IQR pPTC-	IQR pPTC+	Median- differenz	p-Wert (U-Test)	IQR pPTC-	IQR pPTC+	Median- differenz	p-Wert (U-Test)
-19	1,778	1,496	0,283	1	1,623	1,419	0,400	1
-18	1,792	1,324	0,240	5,3560E-31	1,623	1,275	0,148	8,2339E-16
-17	1,768	1,552	0,026	0,8328	1,608	1,457	0,063	0,9998
-16	1,775	1,448	0,017	0,8105	1,632	1,381	0,088	0,9999
-15	1,802	1,425	0,149	1,4114E-10	1,632	1,321	0,061	0,0006
-14	1,771	1,750	0,327	1	1,623	1,537	0,326	1
-13	1,754	1,821	0,153	1	1,613	1,575	0,047	1
-12	1,807	1,511	0,677	0	1,627	1,429	0,597	0
-7	1,781	1,271	0,232	1	2,290	1,577	0,192	1
-6	1,779	1,443	0,336	1,3824E-21	2,352	1,450	0,415	1,2465E-25
-5	1,744	1,800	0,080	0,1794	2,252	2,079	0,309	0,2622
-4	1,760	1,628	0,190	1	2,276	1,625	0,036	0,8242
-3	1,792	1,206	0,203	2,5692E-08	2,372	1,215	0,456	4,1209E-26

## XII. Quantifizierung des NMD-Effektes in Abhängigkeit von Locus und Zelltyp

Motiveigenschaft	Motivanzahl					
	HeLa		HuH7		SH-SY5Y	
	Replikat 1	Replikat 2	Replikat 1	Replikat 2	Replikat 1	Replikat 2
DL						
• Motive mit null Reads ohne PTC	4.347	3.045	13.587	14.862	18.625	23.203
• Motive mit null Reads mit PTC	5.712	4.788	7.722	8.065	13.152	15.261
PL						
• Motive mit null Reads ohne PTC	199	183	2.238	5.311	5.776	5.080
• Motive mit null Reads mit PTC	184	168	422	774	957	971

## XIII. Korrelation der Replikate nach Pearson

Die fettgedruckten Werte stellen die Ergebnisse der DL-Replikatkorrelation dar. Die PL-Ergebnisse verhalten sich invers.

Replikatkorrelation nach Pearson			HeLa		HuH7		SH-SY5Y	
			Replikat 1	Replikat 2	Replikat 1	Replikat 2	Replikat 1	Replikat 2
HeLa	Replikat 1	Korrelation nach Pearson	-	0,299	0,126	0,085	0,069	0,084
		p-Wert	-	1,0499E-206	6,8547E-53	8,5318E-25	3,9686E-17	2,7168E-24
		Probenumfang	-	14738	14738	14738	14738	14738
	Replikat 2	Korrelation nach Pearson	<b>0,259</b>	-	0,14	0,093	0,073	0,082
		p-Wert	<b>2,7060E-47</b>	-	9,1553E-66	1,2251E-29	4,7864E-19	1,3720E-23
		Probenumfang	<b>44235</b>	-	14738	14738	14738	14738
HuH7	Replikat 1	Korrelation nach Pearson	<b>0,044</b>	<b>0,046</b>	-	0,17	0,047	0,058
		p-Wert	<b>2,7631E-20</b>	<b>1,9431E-22</b>	-	2,1478E-17	1,1473E-08	1,4904E-12
		Probenumfang	<b>44235</b>	<b>44235</b>	-	14738	14738	14738
	Replikat 2	Korrelation nach Pearson	<b>0,04</b>	<b>0,043</b>	<b>0,197</b>	-	0,044	0,042
		p-Wert	<b>2,7295E-17</b>	<b>2,4667E-19</b>	<b>4,5553E-15</b>	-	1,0657E-07	4,5897E-07
		Probenumfang	<b>44235</b>	<b>44235</b>	<b>44235</b>	-	14738	14738
SH-SY5Y	Replikat 1	Korrelation nach Pearson	<b>0,006</b>	<b>0,002</b>	<b>0,021</b>	<b>0,016</b>	-	0,202
		p-Wert	<b>0,219</b>	<b>0,626</b>	<b>6,8720E-06</b>	<b>0,001</b>	-	1,1331E-04
		Probenumfang	<b>44235</b>	<b>44235</b>	<b>44235</b>	<b>44235</b>	-	14738
	Replikat 2	Korrelation nach Pearson	-0,013	-0,013	<b>0,014</b>	<b>0,006</b>	<b>0,218</b>	-
		p-Wert	<b>0,008</b>	<b>0,006</b>	<b>0,003</b>	<b>0,181</b>	<b>7,0222E-34</b>	-
		Probenumfang	<b>44235</b>	<b>44235</b>	<b>44235</b>	<b>44235</b>	<b>44235</b>	-

## XIV. Identifizierte SRMs

### • Häufigkeitsverteilung des durchschnittlichen $AI_{sk}$

$AI_{sk}$ -Verteilung	HeLa		HuH7		SH-SY5Y	
	DL	PL	DL	PL	DL	PL
Median	-0,1306	-0,0392	-0,2338	-0,2388	-0,2563	-0,2686
SD	1,0277	0,8034	1,1159	1,0594	1,2149	1,1263
p-Wert (KS-Test)	0,2378	0,4470	0,3193	0,6228	0,8000	0,8506

### • Identifizierte Enhancer- und Silencer-Sequenzen

Die Enhancer- (E) und Silencer-Sequenzen (S) sind alphabetisch geordnet.

HeLa				HuH7				SH-SY5Y			
DL		PL		DL		PL		DL		PL	
Motiv	$AI_{sk}$	Motiv	$AI_{sk}$	Motiv	$AI_{sk}$	Motiv	$AI_{sk}$	Motiv	$AI_{sk}$	Motiv	$AI_{sk}$
(E) AAAAAGTG	2,3975	(E) AAAAGAG	1,6259	(E) AAACACCC	2,9008	(E) AAATATA	2,2607	(E) AACATCAC	2,5937	(E) CCCTACC	2,4683
(E) AAAAAGTC	2,6352	(E) AAAAGAT	1,7742	(E) AAACACCG	3,5776	(E) AACCGAG	2,3534	(E) AATCCGAC	2,5865	(E) CGATTCA	2,6001
(E) AAAAATCG	2,3995	(E) AAAAGCC	1,8362	(E) AAACCTAA	2,4235	(E) AATACAG	2,1716	(E) AATGGCAA	3,0408	(E) CGCAGCC	2,3139
(E) AAAAGAAC	3,5052	(E) AAAATCA	2,0857	(E) AAAGCTAA	2,4577	(E) AATCTAT	3,1007	(E) AATTCAAC	3,1233	(E) CGTCGAA	2,2758
(E) AAAAGACG	3,3305	(E) AAACAAC	1,9553	(E) AATGCGC	2,4105	(E) ACAAAAC	2,2215	(E) ACCAGCCA	2,6384	(E) CTTCAGG	4,4351
(E) AAAAGAGA	2,7142	(E) AAACACA	1,8998	(E) AACAAAGC	2,2956	(E) ACAAGAA	2,3018	(E) ATGTTCCG	2,7345	(E) GAATACA	2,5218
(E) AAAAGCGC	2,1477	(E) AAACCCA	1,9460	(E) AACAAAGC	3,4927	(E) ACACCAC	2,2483	(E) CAACCTAG	2,6479	(E) GACACAA	2,2927
(E) AAAAGGCA	3,1791	(E) AAACCTT	1,9050	(E) AACAAATTA	2,3868	(E) ACATCTGC	2,2311	(E) CAATCTCT	2,4557	(E) TAACCTA	2,5027
(E) AAAAGTAG	2,8947	(E) AAACCTA	2,3362	(E) AACACGCA	2,5254	(E) ACATACA	2,4286	(E) CAGACATA	3,0896	(E) TAATCAA	2,7836
(E) AAAATCCC	2,4427	(E) AAACCTC	1,7032	(E) AACAGCCA	2,5399	(E) ACCATCA	2,2889	(E) CATCTGGA	2,6727	(E) TACTATT	2,5786
(E) AAAATCTG	2,3373	(E) AAACCTT	1,9192	(E) AACCCATA	2,7388	(E) ACTTTTT	2,5863	(E) CATTATTA	2,8501	(E) TCGAACA	2,2579
(E) AAAATGGA	2,8387	(E) AAAGAAT	1,8470	(E) AAGCAGCA	2,9896	(E) ATCAGAC	2,5226	(E) CCCTACTG	2,6716	(E) TCGTAAT	3,0809
(E) AAACAAAG	2,9730	(E) AAAGACA	2,1003	(E) AACTACTA	2,3099	(E) ATTGAAA	2,2020	(E) CCGGGACA	2,6613	(E) TCGTATA	3,0960
(E) AAACAATC	2,1469	(E) AAGCGCC	1,8475	(E) AACTCATA	3,5206	(E) CAACCAT	2,1615	(E) CTATCTCT	2,8237	(E) TGAGGCA	2,5822
(E) AAACATCT	2,4535	(E) AAGGATG	2,0475	(E) AACTGAGC	2,5486	(E) CAAGCCA	2,2177	(E) GATCCGAG	2,6127	(E) TGCATAT	2,6870
(E) AAACATCC	3,0394	(E) AAATBAAT	2,6988	(E) AACTTACA	2,4000	(E) CAAGCCC	2,4017	(E) GCCAAGCG	2,8815	(E) TGCTAAA	2,7105
(E) AAACCGAA	2,6036	(E) AAATAGT	1,7452	(E) AAGACCTC	2,8612	(E) CACAAAA	3,2761	(E) GTACCGAA	2,6820	(E) TGTGCCC	2,3565
(E) AAACCTCC	2,3324	(E) AAATATC	1,6442	(E) AAGACTCG	2,3063	(E) CACGAAT	2,9429	(E) TAATCATA	2,5281	(E) TTCCGAA	3,3146
(E) AAACGAAG	2,2483	(E) AAATCAA	2,0038	(E) AAGAGTAC	2,6245	(E) CAGAAGC	2,2263	(E) TATCTTAA	2,5653	(E) TTGCGCG	2,3767
(E) AAACATATA	3,2856	(E) AAATCCC	1,7676	(E) AAGCTGTA	3,0166	(E) CATGCAA	2,9471	(E) TCGACTGA	3,2486	(E) TTGCGCA	2,4049
(E) AAACATATG	2,2293	(E) AAATTCT	1,6713	(E) AATAATAG	2,2568	(E) CCTTACG	2,8497	(E) TGATCAAA	2,7647	(E) ACCAAAA	2,8423
(E) AAACCTCTA	2,2016	(E) AAATTGT	1,6787	(E) AATCTCA	2,7608	(E) CGAAGAT	2,2797	(E) TTATACAC	2,8876	(E) TTGCGAA	2,5790
(E) AAACCTGCA	2,5157	(E) AAATTTT	1,6842	(E) AATCAAAG	2,5238	(E) CGAATAT	2,2422	(E) AAAGCCTA	2,5585	(E) TTGTCAC	3,0575
(E) AAACCTGCC	2,2926	(E) AACCAAT	2,2421	(E) AATCACTA	3,0759	(E) CGCAAT	2,1673	(E) GAGATTCC	2,5044	(E) AACACAA	2,8496
(E) AAACCTTCA	2,2735	(E) AACCAAC	2,5294	(E) AATCCCAA	3,2748	(E) CGACAGT	2,1338	(E) GGAATAA	3,3308	(E) CGAGACC	3,4308
(E) AAAGATCC	2,5749	(E) AACCACT	2,1431	(E) AATCCCAA	2,4499	(E) CGAGAAT	2,8399	(E) TAACCTAA	2,5525	(E) TAGCGAA	2,3424
(E) AAAGCAAG	2,3063	(E) AAGAAAG	1,7642	(E) AATCCCGC	2,8916	(E) CGAGACC	2,7400	(E) AAGCTATG	2,9593	(E) TGAACAA	2,6851
(E) AAAGCCCA	3,3692	(E) AACGAA	2,3048	(E) AATCGAAG	2,2452	(E) CGAGCAG	2,2157	(E) ACGACCAC	2,5597	(E) TTGCGCC	2,2558
(E) AAAGCCCG	2,7871	(E) AACCRAAT	1,9533	(E) AATGACAC	2,2537	(E) CGAGCGA	2,4413	(E) AGACAGAA	2,7639	(E) CGATAGA	2,8197
(E) AAAGCCTA	2,8014	(E) AACCCAA	2,3164	(E) AATGACCA	2,3221	(E) CGATAAC	2,3726	(E) ATCCCAAC	2,6592	(E) GCACGTA	2,6732
(E) AAAGGACC	2,0990	(E) AAGCAAT	1,6769	(E) AATGAGCC	2,6979	(E) CGCAAAC	2,3808	(E) CGAGCCCA	3,3092	(E) AACCCCC	2,7599
(E) AAAGGATA	3,1090	(E) AACGCAC	1,9608	(E) AATGCAAC	2,4332	(E) CGCGCCG	2,2712	(E) GAAAACTG	3,3368	(E) AATCAAT	2,3932
(E) AAGCTCTC	2,1492	(E) AACTAAA	1,8610	(E) AATGTTTA	2,5166	(E) CGCGGAA	2,6459	(E) GAAAGAAC	2,4773	(E) AATTAAG	2,4601
(E) AAGTTTAC	2,5056	(E) AACTATA	1,7749	(E) AATTAAC	3,0070	(E) CGCGGCC	3,0469	(E) GCGTCAA	2,5916	(E) ACAAAAT	2,3762
(E) AAGTTTCG	2,1881	(E) AACTCAC	1,6996	(E) AATTTCCG	3,0673	(E) CGGAACC	2,5513	(E) TAACACTC	3,1296	(E) AACCCCT	2,3935
(E) AAGTTTAA	2,4057	(E) AACTCCC	1,7319	(E) ACAACCTG	2,8729	(E) CGGAATC	2,1266	(E) TACTTATG	3,1999	(E) ACAGCAA	2,7145
(E) AATAAAC	3,0335	(E) AACTGAG	1,9164	(E) ACAAGAAG	2,3367	(E) CGGCCAG	2,3576	(E) TATCAACC	2,7254	(E) AGCTTAT	2,7291
(E) AATAACAC	2,2261	(E) AAGAAAG	1,8035	(E) ACAGACTC	2,7927	(E) CGGCCCG	3,1240	(E) TCCGAACG	3,8264	(E) CAACATAT	2,4224
(E) AATACTCT	2,4567	(E) AATATCA	1,6406	(E) ACAGACTG	2,8728	(E) CGTATAT	2,1256	(E) TCGAAGCG	2,9178	(E) CACAACA	3,8249
(E) AATAGTA	3,3155	(E) AATCAAA	2,2422	(E) ACAGACCA	2,8869	(E) CGTGATA	2,3606	(E) TCGAGATC	3,1650	(E) CACGGTA	2,8119
(E) AATCCCAA	2,4333	(E) AATCATA	1,7578	(E) ACCAAGCC	2,2908	(E) CTATCGA	2,2947	(E) AACAAAGC	2,5816	(E) CAGCCCT	2,2599
(E) AATCCGAC	2,1416	(E) AATCCGA	1,7509	(E) ACCATTAA	2,3371	(E) CTCCGAG	2,4427	(E) ACTCGAAA	3,1133	(E) CAGTCAA	3,1623
(E) AATGGGCC	2,2305	(E) AATTACA	2,5924	(E) ACCTACGA	2,3601	(E) GAAACAA	3,1987	(E) AGATTACC	3,3264	(E) CATATCT	2,3503
(E) AATTTAC	2,4812	(E) AATTATA	1,9334	(E) ACCTATTA	2,3777	(E) GAAACTC	2,1613	(E) GGATCCTG	6,6129	(E) CCGGCGA	2,3712
(E) AATTTCCC	2,6994	(E) AATTTTC	1,9357	(E) ACCTGAAA	2,4812	(E) GACCAAA	2,1428	(E) TAGAAGAG	2,8162	(E) CGGCCAT	2,7803
(E) AATTTCCG	2,5523	(E) ACACATC	2,1013	(E) ACGAAGAA	2,6143	(E) GATTACT	2,1241	(E) TCGCTATA	3,0189	(E) CTCCCAA	2,3267
(E) AATTTGCA	2,6388	(E) ACAGAAAT	1,9712	(E) ACGAAATA	3,2329	(E) TAACCTGC	2,1274	(E) AGAACCAAC	2,7966	(E) GAAAACT	2,5342
(E) AATTTGGA	2,5222	(E) ACATAAA	1,9385	(E) ACGACTCA	3,2675	(E) TACAAC	2,6044	(E) TCGACTCG	2,7555	(E) GACTAAA	2,3878
(E) AATTTTCC	2,5595	(E) ACATAAAC	1,6202	(E) ACGAGACC	2,6106	(E) TACCAAC	2,3002	(E) TCGGTCCA	2,6261	(E) GCGAAAA	2,3672
(E) AACRAAAC	2,1772	(E) ACATGAA	1,7716	(E) ACGCAACG	2,4239	(E) TAGCGAA	2,2753	(E) TTGCACCA	2,4560	(E) GCGAAAA	2,5125
(E) AACRAACA	4,3500	(E) ACATTAT	1,7055	(E) ACGCACCA	2,3604	(E) TATAACA	2,2499	(E) AAATGCAA	3,0957	(E) TAAACTT	2,5045
(E) AACACATA	2,4744	(E) ACATTTA	1,8967	(E) ACGGAGTA	2,9086	(E) TCAAACT	2,2003	(E) AACAAATG	2,7958	(E) TAACATC	2,4696
(E) AACACCCA	2,1829	(E) ACCCAAA	1,6514	(E) ACGTAGCA	2,4671	(E) TCACTCT	2,6892	(E) AACCTGCC	2,8156	(E) TAATACT	2,6607
(E) AACACCTG	2,1551	(E) ACCTATA	1,6691	(E) ACGTCTCA	2,2781	(E) TCCGGCT	2,6183	(E) AACTCTAA	2,9204	(E) TAATCAC	2,3972
(E) AACACGAA	2,7719	(E) ACGGAAT	1,9285	(E) ACTCGAAA	2,3731	(E) TCGACGA	2,4291	(E) AAGCTCGA	3,9846	(E) TACTAAC	2,4226
(E) AACACTCC	2,4151	(E) ACGAAGT	1,7005	(E) AGAAACCA	3,3035	(E) TCGTAAG	2,8403	(E) AATACTGG	2,7813	(E) TAGAACT	2,7956
(E) AACACTGA	2,6074	(E) ACGCATC	1,6454	(E) AGAACGAA	2,3050	(E) TCGTACT	2,7627	(E) AATTACTA	2,7798	(E) TATGACC	2,2758

(E) AACATCTA	2,1401	(E) ACGTAA	1,8743	(E) AGAAGATA	2,9431	(E) TGAACAA	2,3717	(E) ACACCTAC	2,4816	(E) TCAGCCA	2,8802
(E) AACATCTG	2,0706	(E) ACTTAA	2,2331	(E) AGACCCTA	3,3952	(E) TGAATAC	2,4018	(E) AGAATCGA	2,6577	(E) TCCGGCC	3,6751
(E) AACATTAC	2,1424	(E) AGAAAAC	2,0080	(E) AGACGCAA	2,9195	(E) TGACCAT	2,2309	(E) CACTATAA	2,7327	(E) TCGCCCT	2,5301
(E) AACCAAAC	2,2690	(E) AGAAAT	2,0687	(E) AGAGAAAG	2,4594	(E) TGCAACC	2,8189	(E) CATAAATA	2,5165	(E) TCTGGAA	2,3356
(E) AACCAACG	2,2727	(E) AGAAATT	1,7474	(E) AGATACCA	2,3856	(E) TGCAAGC	2,1681	(E) CCATGACC	2,7605	(E) TGACAGC	2,3082
(E) AACCAATA	2,2769	(E) AGAAACCA	1,6868	(E) AGATTACC	2,2640	(E) TGCAGCA	2,6161	(E) CCCCATA	2,7322	(E) TGCCCAA	2,4144
(E) AACCAAGC	3,2595	(E) AGAACTA	1,6512	(E) AGATTCCC	2,3286	(E) TGCAGCT	2,6899	(E) CCTAGATC	2,4824	(E) TTATATA	2,4288
(E) AACCATAG	2,0576	(E) AGACAAA	2,2799	(E) AGCAATC	2,2445	(E) TGCAGCG	2,8351	(E) CGAGACCC	2,4796	(E) CAAACGA	2,7033
(E) AACCATCA	2,4228	(E) AGACCAT	1,6267	(E) AGGAAACC	2,6611	(E) TGCAGGAG	3,3841	(E) GAAGGAAA	2,8047	(E) TCCGATA	2,5741
(E) AACCCAGA	3,6511	(E) AGACGAA	2,4121	(E) AGTCGATC	2,2570	(E) TGGAAAA	2,4232	(E) GACCGCAA	2,5855	(E) AAAACAA	2,5239
(E) AACCCCCC	2,2697	(E) AGACTCT	3,5346	(E) ATACACAC	2,6058	(E) TGGCACT	2,4133	(E) GACGGAGA	2,9644	(E) TAAACTA	2,6530
(E) AACCCCTG	2,6657	(E) AGAGAA	2,1949	(E) ATACAGAA	3,0073	(E) TTACCCC	3,0431	(E) GATGAAC	2,6077	(E) TACACCA	2,3204
(E) AACCCGAC	2,4639	(E) AGAGACA	1,7147	(E) ATACCGAA	2,3545	(E) TTCAATT	2,3249	(E) GCAAGTAC	2,8806	(E) TTCCGAC	2,2633
(E) AACCTTAC	2,0969	(E) AGAGACC	1,7081	(E) ATATCTAA	3,1878	(E) TTCCAAC	2,3654	(E) TAGAGCAC	2,5218	(S) ATTGGGT	-2,5528
(E) AACCGATC	2,0593	(E) AGAGCAC	2,1398	(E) ATATGGAC	2,2655	(E) TTCCAC	2,9624	(E) TATATATA	2,7069	(S) GGGAGGT	-4,3584
(E) AACCTTCC	2,1370	(E) AGATAAA	1,6933	(E) ATCAAACG	2,5516	(E) TTGCATC	2,9269	(E) TCAAGCTA	2,7726	(S) ACGGGGT	-2,2965
(E) AACGAAAG	3,2367	(E) AGATAAT	1,6438	(E) ATCCACCA	2,8462	(E) TTGCATG	2,1561	(E) TCAATATG	2,8433	(S) AGGGTTG	-2,4166
(E) AACGAATG	2,4057	(E) AGATCCA	2,0957	(E) ATCGAGAG	2,2826	(E) TTGCGCA	2,4486	(E) TCAGAGCA	2,8882	(S) AGTGGGG	-2,5812
(E) AACGACCG	2,5247	(E) AGACATA	1,7419	(E) ATCTTCGA	2,2837	(E) TTGCTGT	2,9112	(E) TCGAGATG	2,8691	(S) ATTCGGG	-2,4915
(E) AACGAGAG	2,2169	(E) AGCGAAA	1,9566	(E) ATCTTTCA	2,3513	(E) TTCTGAT	2,1189	(E) TCGCTGGA	3,0168	(S) GGGCGGG	-4,4412
(E) AACGAGGG	2,1436	(E) AGTTAAT	1,9050	(E) ATTAAGCC	2,2333	(E) TTGCAGG	2,4672	(E) TCGGGAGC	2,6356	(S) GGGGATT	-3,0921
(E) AACGCAAC	2,3900	(E) ATATATA	2,0268	(E) ATTACTAG	2,5634	(E) TTGCCAC	2,2263	(E) TGACCAAA	2,4334	(S) GGGGCGC	-2,6083
(E) AACGCCCC	2,3302	(E) ATATATT	1,8112	(E) CAAACAAC	3,2950	(E) TTGCCAG	3,2282	(E) TGATATCA	2,9268	(S) GGGGCTT	-2,5538
(E) AACGCTGA	2,6392	(E) ATCAAT	2,2647	(E) CAAGAAGC	3,1380	(E) TTGCCGT	2,2454	(E) TTGGCGTA	2,7272	(S) GGGTAGT	-2,4891
(E) AACGTACC	2,6806	(E) ATCTATA	1,8839	(E) CACAGAAA	2,2688	(E) TTGTGAA	2,2415	(E) AAAGGATA	2,5012	(S) GGGTGGT	-2,3239
(E) AACTAABC	3,0916	(E) ATTAATA	2,8074	(E) CACTGAAA	2,4536	(E) TTCCCAA	2,3769	(E) AACCCAGA	2,5027	(S) GGTGTCT	-3,4711
(E) AACTAAGA	2,4570	(E) ATTAGCC	1,6307	(E) CATACACA	2,8299	(E) ACATAAA	2,8873	(E) AATCACC	2,6230	(S) GTCTTGT	-2,4444
(E) AACTAAGC	2,0714	(E) CAAACAA	2,8898	(E) CATACATG	2,4241	(E) CACGGAA	2,1197	(E) GAACGAGA	2,6248	(S) GTGCAGA	-2,6182
(E) AACTACCG	2,3876	(E) CAAAACG	1,7552	(E) CATCATAA	2,5089	(E) CAGCAAC	2,7005	(E) GAAGATAA	2,4514	(S) GTGGTCG	-2,5892
(E) AACTACGG	2,3971	(E) CAAAATC	2,0061	(E) CCAAGAG	2,2454	(E) CGAAGAA	2,8648	(E) GCCGAAAG	2,5562	(S) TGCCCGG	-2,5121
(E) AACTCAAC	2,3214	(E) CAAAATC	1,7230	(E) CCACTCAA	2,3397	(E) CGATACT	2,6351	(E) TCGACAAA	2,8382	(S) TGTGTCA	-2,3565
(E) AACTCACA	2,7555	(E) CAAACCA	2,4215	(E) CCTGAGA	2,2590	(E) CGGCCCA	2,3426	(E) TCGGACAA	3,1004	(S) ATGGGGT	-2,7377
(E) AACTCACC	3,6106	(E) CAAACGA	2,2172	(E) CCGAATA	2,6017	(E) GACGAAA	2,8330	(E) AAAAGGCC	2,7222	(S) GCCCGGT	-3,0820
(E) AACTCCGA	2,1193	(E) CAAAGAT	2,1139	(E) CCGCGCA	2,3067	(E) TACCCCC	2,8681	(E) AAACCTTA	2,9129	(S) GGGACTC	-3,5848
(E) AACTGTAC	2,0989	(E) CAAATAC	1,9603	(E) CCTAATA	2,2613	(E) TACGCAG	2,3494	(E) AAACGAAC	2,9875	(S) GGTACTG	-2,5286
(E) AAGAACA	2,2015	(E) CAAATCT	1,8430	(E) CCTAGAAA	2,2711	(E) TCAAAAT	2,3395	(E) AACGAGGA	2,5102	(S) GTGGGCC	-2,6380
(E) AAGAAAGG	2,4729	(E) CAACACT	3,2004	(E) CGATTCTA	2,2419	(E) TCCGGAG	2,3860	(E) AACGATTA	2,7116	(S) GTGGGCT	-2,5347
(E) AAGAACCA	3,1526	(E) CAACGCC	1,9045	(E) CGGAGACA	2,8258	(E) TCGCTCC	2,6011	(E) AACGTCCA	2,8371	(S) AAGGGTG	-3,0882
(E) AAGAACTA	2,2069	(E) CAACGCA	2,0753	(E) CGGCGATG	2,3549	(E) TTCACCC	2,3255	(E) AAGAGACC	2,4527	(S) AAGTGCA	-3,4490
(E) AAGAAATC	2,5855	(E) CAACCTC	2,2208	(E) CTACTCTC	2,2803	(E) TTGAGCA	2,1884	(E) AAGATTAC	3,0272	(S) ACCCGGT	-2,4227
(E) AAGAAATG	2,4151	(E) CAACCTT	1,6804	(E) CTCACCAA	2,6469	(E) TTGCCCC	2,9012	(E) AAGCTGGA	2,6429	(S) ACCTCGC	-3,9117
(E) AAGAATTA	2,2697	(E) CAAGAAT	2,3438	(E) GAAAACCC	2,4497	(E) TTGCCCT	2,3484	(E) AAGGCAAC	2,5201	(S) ACGAGTA	-2,4610
(E) AAGACAA	2,4275	(E) CAAGACA	2,3980	(E) GAAACCGC	2,3652	(E) TTGGCCC	2,1625	(E) AATCGACC	2,8955	(S) ACTCGGC	-3,1389
(E) AAGACACC	2,4190	(E) CAATCAA	3,2185	(E) GAAATTCG	2,3538	(E) AACGAAC	2,6804	(E) AATGAATC	2,4388	(S) ACTCGGG	-2,4576
(E) AAGACTTA	2,7119	(E) CAATCCA	1,7478	(E) GAACAGAA	2,5334	(E) AAGTGAA	2,2932	(E) ACGAAATC	3,3925	(S) ACTGTCG	-2,6806
(E) AAGAGTAG	3,0796	(E) CACACCA	1,7912	(E) GAACCTTA	3,2231	(E) CAACCGA	2,3426	(E) AGCTAAAA	2,8955	(S) AGGACGG	-2,9234
(E) AAGAGTTC	2,1584	(E) CACACCC	1,6945	(E) GAATATGA	2,3253	(E) CACGACT	2,2556	(E) CGAAGATA	3,8056	(S) AGGAGCC	-2,9528
(E) AAGATTGC	2,0992	(E) CACCCCA	1,7087	(E) GAATCCTA	2,4335	(E) CCGCAAA	2,4921	(E) CGACATCA	2,5311	(S) AGGCGCG	-2,8952
(E) AAGCAATA	2,8422	(E) CACCTTG	2,0877	(E) GAATCCCA	2,3647	(E) CGACCCA	2,9035	(E) CGACCAAC	2,6880	(S) AGGGCGG	-2,4957
(E) AAGCACAC	2,1688	(E) CACGAAC	2,5153	(E) GACCATCA	2,2865	(E) CGATAGA	2,6729	(E) CGATAAGA	2,7169	(S) AGGGGTG	-2,4424
(E) AAGCACTC	2,0908	(E) CACGATG	1,9630	(E) GACGACTG	2,3730	(E) GAATTAT	2,3625	(E) CGTAGAAA	3,0197	(S) AGGTTCT	-2,4678
(E) AAGCCCCG	2,8974	(E) CACGGAA	2,4705	(E) GAGAATTG	3,0002	(E) GCACGTA	2,4005	(E) GAACCCGA	2,8953	(S) AGTAGTG	-2,6155
(E) AAGCCTGA	2,5118	(E) CACTAAC	1,8300	(E) GAGATACA	2,4714	(E) GCACCTC	2,2029	(E) GACGCAGA	2,9459	(S) AGTCAGG	-3,2822
(E) AAGCGAAA	2,2163	(E) CACTATA	2,4615	(E) GAGATTAG	3,0127	(E) TAGAAAA	2,2893	(E) GAGTTCGA	3,7461	(S) AGTCTGT	-4,1333
(E) AAGCGAGC	2,4642	(E) CACTCCC	2,1858	(E) GAGTCAAC	2,2502	(E) TCACATC	2,2366	(E) GATCAAAA	3,0789	(S) AGTGGTC	-3,0702
(E) AAGCTACC	2,2410	(E) CAGAACA	2,7000	(E) GAGTGACA	2,3487	(E) TCGCGTA	2,3814	(E) GATTATAA	2,9350	(S) AGTTGGG	-3,3638
(E) AAGGACAA	2,9416	(E) CAGAACC	1,7924	(E) GATAACAG	2,3549	(E) TTGCGCG	3,2304	(E) TAAATTC	3,1655	(S) ATATTTT	-2,4454
(E) AAGAGCTA	2,3514	(E) CAGAACT	2,4764	(E) GATACCTA	2,2491	(E) TTGCGGA	3,2033	(E) TAACACAC	2,5093	(S) ATGGCGC	-2,6965
(E) AATACACC	2,8644	(E) CAGACCG	1,7862	(E) GATTCCAA	2,5080	(E) TTGTAGC	2,4966	(E) TAAGAGCA	2,6219	(S) ATGGCTT	-2,3759
(E) AATAACGG	2,3989	(E) CAGCAAT	1,8873	(E) GCGACTCG	3,0455	(S) AAGGGGC	-3,2506	(E) TACAAAAC	2,5263	(S) ATGTGTG	-2,4931
(E) AATAACTG	2,3028	(E) CAGCGTA	1,7685	(E) GCCCCTAA	2,9573	(S) AGGGGGT	-3,3424	(E) TACGAACA	2,9286	(S) ATTGGCG	-3,1677
(E) AATACATC	2,5589	(E) CAGCCTC	1,7226	(E) GCCGATAA	2,4626	(S) AGGGGTA	-2,5442	(E) TCCGCCCA	2,8864	(S) CATAGTG	-2,5371
(E) AATACGAC	2,5064	(E) CAGCTAC	1,7151	(E) GCGAACAA	2,8023	(S) AGGGTCT	-2,2948	(E) TCGGTGTG	3,3450	(S) CATCTGG	-2,3117
(E) AATAGACC	2,7723	(E) CAGTAAC	1,6150	(E) GCGAATCA	2,2837	(S) AGTGCGT	-2,3261	(E) TTGGCGGA	2,5713	(S) CATCTTT	-2,8585
(E) AATAGCAA	2,5301	(E) CATCCAA	1,6758	(E) GCGAATCG	2,5501	(S) ATGGGGC	-2,7541	(S) CGCGGCG	-2,7226	(S) CCGGCTT	-2,4569
(E) AATAGGAA	2,2431	(E) CATCCCA	1,6094	(E) GCGATTG	3,1979	(S) ATTGGGT	-2,3815	(S) CTTCGTG	-3,4193	(S) CCTCGGG	-2,8226
(E) AATAGGCA	2,0721	(E) CCAAAAT	2,2335	(E) GCGTCATA	2,3580	(S) CAGGGGA	-2,2170	(S) GGGGGTCG	-4,3496	(S) CGCTGTT	-2,9820
(E) AATAGTGA	2,2068	(E) CCAAGAA	1,6894	(E) GCGTCCCA	2,4074	(S) CCACGGT	-3,0911	(S) GTTGGGTA	-3,6208	(S) CTCAAC	-3,1134
(E) AATATACA	2,8034	(E) CCAACGA	2,3727	(E) GCTGAAA	2,6746	(S) CCTTGCT	-2,4385	(S) AACGGGGA	-2,5310	(S) CTCTAGC	-2,7391
(E) AATATCAG	2,5318	(E) CCACAAC	2,0209	(E) GGATCCTG	3,9419	(S) GAGGGCA	-2,2185	(S) AGCGGGGA	-3,4727	(S) CTCTGAC	-2,3025
(E) AATCAATA	3,0441	(E) CCACAC	1,7365	(E) GTATCTAA	2,4170	(S) GAGGTGC	-3,0833	(S) AGGGGGCG	-2,4446	(S) CTTCGGG	-2,9809
(E) AATCACCG	2,5412	(E) CCACGAT	1,6133	(E) TAAATCTT	2,9147	(S) GCGTGGC	-2,8249	(S) AGGGGGTC	-4,2628	(S) GACATCT	-3,8629
(E) AATCAGCA	2,1636	(E) CCACGCC	1,8115	(E) TAAATGGA	2,4277	(S) GCTAAT	-2,1655	(S) AGTGGGG	-3,9371	(S) GATCCTG	-2,8513
(E) AATCATCA	3,0101	(E) CCACATAC	1,6503	(E) TAACGAGA	2,2856	(S) GGACGGT	-3,1303	(S) CACCGGG	-2,7996	(S) GATCGGC	-2,3477
(E) AATCATCC	2,1047	(E) CCATAAA	1,9605	(E) TAAGAAAC	2,9303	(S) GGCACGT	-2,3968	(S) CCCGGGG	-3,7486	(S) GCAGAGG	-2,4388
(E) AATCCATC	2,0973	(E) CCACTAA	1,9625	(E) TAAGCAGC	2,5270	(S) GGGAGGT	-4,0621	(S) CCGCTTG	-2,4445	(S) CGCGCC	-4,1376
(E) AATCCCAC	2,7587	(E) CCAAAAA	1,8726	(E) TAATGACA	2,4178	(S) GGGGTCT	-2,5188	(S) CCTCGGG	-2,4752	(S) GCGGCA	-3,0570
(E) AATCCCTC	2,1513	(E) CCAACAT	2,2387	(E) TACAACCC	2,6636	(S) GGTACCG	-3,2677	(S) CGCAGCTG	-2,6577	(S) GCCTGGG	-4,8005
(E) AATCCTCA	2,3310	(E) CCCACCC	1,9678	(E) TACACCAG	2,5401	(S) GGTACGG	-3,0981	(S) CGCAGGGA	-2,6176	(S) GCGACTG	-2,7416
(E) AATCCTTA	2,6806	(E) CCACTA	1,9306	(E) TACCACAA	2,7210	(S) GGTAGGG	-2,6134	(S) CGCGGCTC	-2,9573	(S) CGCAAC	-2,4514
(E) AATCGAAA	3,5345	(E) CCCAGAT	1,7880	(E) TACCACCG	2,8612	(S) GGTCTCT	-2,4178	(S) CGCGGGTC	-2,6626	(S) GCGGCGC	-4,4260

(E) AATCGAAC	2,4994	(E) CCCCCA	2,3159	(E) TACCCAAC	3,0511	(S) GGTCTCT	-2,7293	(S) CGCGTGGG	-3,1430	(S) GCGGGAG	-2,5108
(E) AATCGGAG	2,3247	(E) CCGGACC	2,1864	(E) TACCTGCG	2,3061	(S) GGTGGTC	-2,2398	(S) CGCGTGTG	-2,6088	(S) GCGGGCT	-2,4632
(E) AATCTCAG	2,5556	(E) CCGGAC	1,6376	(E) TACCTCAA	2,2903	(S) GGTGTCC	-2,3549	(S) CGCGGGG	-3,2541	(S) GCGGGGA	-3,0620
(E) AATCTCCA	3,1097	(E) CCGGGAG	1,6322	(E) TAGAACTC	2,3796	(S) GTCACTG	-2,1215	(S) CGGGGGTA	-3,4683	(S) GCGGGTG	-3,1195
(E) AATCTCCC	2,7681	(E) CCGGAACA	1,9921	(E) TAGAAGAG	2,2833	(S) GTGGGTA	-2,9260	(S) CGGGTGGG	-2,7963	(S) GCTCGAG	-4,0556
(E) AATCTCTCA	3,0640	(E) CCGGCTA	2,0405	(E) TAGAATCA	2,5197	(S) GTTGGGT	-2,5193	(S) CGTCGGGA	-2,9085	(S) GCTGGCT	-3,2080
(E) AATCTTTA	2,1630	(E) CCGGCAG	1,8481	(E) TAGAGATC	2,2343	(S) GTTGTGC	-2,1487	(S) CGTCGGGG	-3,2966	(S) GCTTCTG	-2,4687
(E) AATGAACC	2,0620	(E) CCGTCCC	2,2736	(E) TAGATATC	2,6239	(S) TAGGGCA	-2,1401	(S) CGTGGGGC	-2,5983	(S) GGAGGCC	-2,3628
(E) AATGATAA	3,0186	(E) CCTAGAC	1,9854	(E) TAGATTGA	2,5302	(S) TAGGGGT	-2,3194	(S) CGTGGTGC	-2,7021	(S) GGAGGGG	-2,3390
(E) AATGATAC	2,5179	(E) CGAAAAC	2,0592	(E) TAGATTGT	2,4708	(S) AAAGAAG	-2,5589	(S) CGTGGGG	-3,3471	(S) GGCACTG	-2,4802
(E) AATGATAG	2,1814	(E) CGAAACC	2,1541	(E) TAGCCAGC	2,5456	(S) AAGGCTA	-2,8923	(S) CTCTGGGC	-2,7964	(S) GGCAATG	-3,2045
(E) AATGATTA	2,3706	(E) CGAAATG	1,8783	(E) TAGCTACG	2,5920	(S) AAGGGAG	-2,9191	(S) CTCTTGGC	-3,2633	(S) GGCAATG	-2,2907
(E) AATGCATC	2,0828	(E) CGAACCA	2,0307	(E) TAGGAGAC	2,3826	(S) AAGGGCG	-2,4049	(S) CTGCTGGC	-2,4374	(S) GGGGGGG	-3,9025
(E) AATGCTAA	3,0012	(E) CGAACCC	1,6193	(E) TAGTGATG	2,2598	(S) AAGGGTT	-2,2797	(S) CTGGGGCG	-2,7510	(S) GGGACAG	-2,4010
(E) AATGCTAG	2,2610	(E) CGAACC	2,2639	(E) TATAATTA	2,6688	(S) AAGTTCT	-3,0640	(S) CTGGGGCG	-3,6206	(S) GGGACGG	-2,2611
(E) AATGGAAC	2,2030	(E) CGAAGCT	1,6219	(E) TATCAACA	2,4791	(S) ACATGCG	-2,1380	(S) GAGGGGGG	-3,8719	(S) GGGATCG	-3,0749
(E) AATTAATC	2,3620	(E) CGAAGAA	1,9449	(E) TCAAACGA	2,4487	(S) ACCGAGT	-2,1306	(S) GCTCGTGC	-2,4493	(S) GGGATTT	-4,0914
(E) AATTACAC	2,8919	(E) CGAATAC	2,0687	(E) TCAAACCT	2,4082	(S) ACCGGTC	-2,5303	(S) GGGCCGGG	-3,7383	(S) GGGCGCT	-3,1801
(E) AATTACCA	3,0913	(E) CGAATCC	2,1329	(E) TCACATCC	2,4743	(S) ACGATAC	-2,1511	(S) GGGCGGCG	-3,2237	(S) GGGCGGT	-2,3311
(E) AATTACTC	2,2493	(E) CGACAAA	2,2486	(E) TCAGAAATC	2,2513	(S) ACGCATT	-2,2873	(S) GGGCGGGC	-3,9867	(S) GGGGAAG	-3,2955
(E) AATTATAG	2,2812	(E) CGACAGA	1,9096	(E) TCAGGAAG	2,3538	(S) ACGCGCG	-4,3336	(S) GGGGAGGA	-3,5178	(S) GGGGAGT	-2,3543
(E) AATTATGA	2,7148	(E) CGACCCG	2,0880	(E) TCCGAATC	2,9429	(S) ACGGGAA	-2,4505	(S) GGGGGCAA	-3,8801	(S) GGGGCAG	-2,4374
(E) AATTATTA	2,1030	(E) CGACCTA	2,3044	(E) TCCGACAA	2,3545	(S) ACGGGGC	-2,7022	(S) GGGGGCTG	-3,4792	(S) GGGGCGC	-2,3234
(E) AATTCAAA	2,7731	(E) CGACGCA	2,2051	(E) TCCGAGAA	2,3008	(S) ACGGGGT	-3,1740	(S) GGGGGGTA	-2,7027	(S) GGGGGCG	-2,3021
(E) AATTCAAG	2,0904	(E) CGACTAT	1,8411	(E) TCCTCCAC	2,3512	(S) ACTCTGT	-2,6799	(S) GGGGGTAG	-3,2752	(S) GGGGGAC	-2,4420
(E) AATTCAACC	2,6298	(E) CGACTCT	1,6992	(E) TCGAATCC	2,5548	(S) ACTGCTG	-2,2092	(S) GGTGGGGG	-3,8046	(S) GGGTAAA	-3,1848
(E) AATTCAAGC	2,1950	(E) CGAGACT	2,7921	(E) TCGACTTA	2,2345	(S) ACTGGGT	-2,4286	(S) GGTGGGAC	-2,7837	(S) GGGTCAAT	-2,8441
(E) AATTCTCTA	2,4136	(E) CGAGATG	1,9988	(E) TCGAGGAA	2,6775	(S) ACTTATT	-3,0405	(S) GGTGGGGG	-3,7730	(S) GGGTCGC	-2,9698
(E) AATTCTGAA	3,3584	(E) CGAGCAA	1,9270	(E) TCGAGGCA	2,5944	(S) ACTTTCT	-2,3233	(S) GGTGTGTC	-2,7100	(S) GGGTCGG	-2,5051
(E) AATTGACA	2,3846	(E) CGAGACT	1,6502	(E) TCGAGGCC	2,4502	(S) AGAGGCT	-2,5696	(S) GTACGGGG	-4,1152	(S) GGGTGGC	-3,1124
(E) AATTGGAG	2,0602	(E) CGAGCCC	2,0983	(E) TCGATGGA	2,5561	(S) AGAGGTT	-2,2529	(S) GTCCGGGG	-4,4937	(S) GGGTGTG	-2,8245
(E) AATTTCAA	2,4927	(E) CGAGCCG	1,7770	(E) TCGATGGC	2,2966	(S) AGACATC	-3,0792	(S) GTCTGTGC	-2,6532	(S) GGGTGTG	-2,2555
(E) AATTTCGT	2,3151	(E) CGATAAT	1,8827	(E) TCGCAGGC	3,0007	(S) AGCAGTG	-2,4013	(S) GTCTGGGG	-3,8298	(S) GGGTCTT	-2,8633
(E) AATTGGCC	3,0502	(E) CGATACC	2,4809	(E) TCGCGAGC	2,9014	(S) AGCGATC	-2,2355	(S) GTGGGGCG	-2,8309	(S) GGTCCCG	-2,7161
(E) AATTTCCTA	2,6625	(E) CGATACT	2,5123	(E) TCGCGGCC	2,7305	(S) AGCGCGG	-3,1318	(S) GTGGGTGG	-2,8151	(S) GGTCCCG	-5,2492
(E) ACAAGGGA	2,3366	(E) CGATCAG	2,0949	(E) TCGGAAAA	2,2387	(S) AGGAGCG	-2,9422	(S) GTTAGGGG	-4,0253	(S) GGTCCGG	-2,8341
(E) ACAAGATG	2,2455	(E) CGATGAT	1,6862	(E) TCGGAGGC	2,2490	(S) AGGCTGG	-3,4202	(S) GTTCGGGG	-3,0988	(S) GGTCCCT	-3,3584
(E) ACAATAAA	3,2057	(E) CGATGTT	2,6407	(E) TCGGATTG	2,5749	(S) AGGGAGG	-2,3920	(S) GTTGGGAG	-3,6049	(S) GGTCCGG	-2,5982
(E) ACAATAAG	2,5652	(E) CGCAAAA	3,0757	(E) TCGCGCTC	2,6115	(S) AGGGATT	-3,7482	(S) GTTGGGGA	-3,3692	(S) GGTGGAT	-3,6454
(E) ACAATACA	2,6721	(E) CGCATAC	1,7104	(E) TCGGGTAG	2,8417	(S) AGGGGCA	-2,8145	(S) TGCGTTTA	-2,4331	(S) GGTGGCT	-2,3369
(E) ACAATTTG	2,2637	(E) CGCCACA	2,0907	(E) TCGTCCAA	3,5494	(S) AGGGGGC	-2,3135	(S) TGTGGGG	-3,8117	(S) GGTTCAG	-3,4181
(E) ACAACCTC	2,2068	(E) CGCGATA	1,6847	(E) TCGTCTGT	2,7420	(S) AGGGTTG	-2,4524	(S) AAGGGTGG	-3,5764	(S) GGTTCAG	-2,7634
(E) ACAAGAGA	2,7891	(E) CGCGACC	2,0302	(E) TCGTTATA	2,7401	(S) AGTCGGA	-2,4253	(S) AAGTGGGG	-2,6193	(S) GTACCGG	-3,9891
(E) ACAAGCTA	2,3006	(E) CGGCCAA	2,3559	(E) TCTAATTC	2,3085	(S) AGTCGTC	-2,7804	(S) ACGGGGCG	-3,7347	(S) GTCCGCT	-2,3952
(E) ACAAGGAA	2,3614	(E) CGGCCAC	2,1251	(E) TCTACATA	2,4575	(S) AGTGATT	-2,2221	(S) AGGGGGAA	-3,3144	(S) GTCCGTC	-2,3363
(E) ACAATCCC	2,7014	(E) CGGCCCA	2,9129	(E) TCTGAGAC	2,6024	(S) AGTGGGC	-2,5935	(S) AGTCGGGG	-3,3543	(S) GTCTGTC	-2,8171
(E) ACAATGCC	2,5624	(E) CGGCCCC	2,6171	(E) TGAACAA	3,8517	(S) AGTGGGG	-2,7036	(S) CGGGGGTC	-3,4734	(S) GTCTGCT	-2,7558
(E) ACACACAA	2,3226	(E) CGGCTCC	1,7660	(E) TGAAGAA	2,6572	(S) AGTTGCA	-2,9498	(S) CGAAGGGG	-3,6930	(S) GTCTTGG	-2,3158
(E) ACACACCC	2,3869	(E) CGGCTAC	1,6355	(E) TGACGACC	2,2805	(S) AGTTCTT	-2,6712	(S) CGCGGGCG	-3,6850	(S) GTCCGGG	-3,4452
(E) ACACATAA	2,3918	(E) CGTATA	1,9224	(E) TGAGAACCA	2,9004	(S) AGTTGGT	-2,9335	(S) CTCTCGGG	-2,5231	(S) GTGGGAT	-2,4046
(E) ACACCCCTA	2,7794	(E) CGTACAA	1,6212	(E) TGATTTC	2,6147	(S) ATATGAC	-2,1349	(S) CTGGGGTA	-4,0142	(S) GTGGCCG	-2,4368
(E) ACACATACC	3,2784	(E) CGTCCCG	1,8493	(E) TGGAAACG	2,2717	(S) ATCCCGG	-2,1973	(S) CTTTGGG	-3,4969	(S) GTGGGAG	-2,8716
(E) ACAGAACCA	2,1350	(E) CTATAA	3,0230	(E) TGGCCCCA	2,6316	(S) ATCCCGT	-2,5892	(S) GCATGGTC	-2,5701	(S) GTGGGGT	-2,6557
(E) ACAGACTA	2,1213	(E) CTATCTG	1,9832	(E) TGGGCACG	3,3006	(S) ATCCGGG	-2,5250	(S) GGGCACGG	-4,5620	(S) GTGTCTG	-2,6848
(E) ACAGCACCC	2,1385	(E) CTCAACC	2,0951	(E) TGTAACAA	3,1180	(S) ATCCGGT	-2,1322	(S) GGGCAGTG	-3,4473	(S) GTGTCCG	-2,2823
(E) ACAGCCCG	2,2729	(E) CTCACGA	1,8106	(E) TTACCCTA	2,3296	(S) ATGCCAT	-2,2742	(S) GGGCTGTG	-3,3000	(S) GTGTCTC	-2,4627
(E) ACAGGCCA	2,1678	(E) CTCACCT	1,9745	(E) TTGCGAGG	2,4602	(S) ATGGCGG	-2,4088	(S) GGGGAGTG	-3,7387	(S) GTTAAGT	-3,0435
(E) ACATAAAA	2,3660	(E) CTCTGAA	1,7761	(E) TTCTACCA	2,7044	(S) ATGGGGC	-2,7705	(S) GGGGGAAG	-2,6351	(S) GTTCGCT	-2,3136
(E) ACATAACC	2,6333	(E) CTTGACC	1,7115	(E) TTGCAGCC	2,3424	(S) ATGGGCT	-2,2987	(S) GGGGGGAG	-4,0027	(S) GTTCTGC	-2,8407
(E) ACATATCG	2,1584	(E) GAAATAT	1,6279	(E) TTGAGAGC	2,2434	(S) ATGTGGC	-2,3778	(S) GGGGGTAC	-2,7711	(S) GTTCTTT	-2,3356
(E) ACATGACC	2,1770	(E) GAACCTA	1,7456	(E) TTTGAAAC	2,9200	(S) ATGTGGT	-2,3841	(S) GGTGGGCC	-3,9848	(S) GTTGGTT	-3,3191
(E) ACATTACC	3,5097	(E) GACAAAT	1,7109	(E) TTTTGACC	2,6014	(S) ATTCGGG	-3,4259	(S) GTGGGTTT	-3,3075	(S) GTTGTGT	-2,8759
(E) ACATTCAA	2,0926	(E) GACATCT	1,8765	(E) AAATACTC	2,8730	(S) ATTGCTT	-2,3791	(S) TCCACTGC	-2,7180	(S) GTTTTCG	-2,3320
(E) ACCAACTC	2,0554	(E) GACCCAA	1,8025	(E) AAATTTCC	2,2601	(S) ATTGTCC	-2,4025	(S) TGCTGGGG	-3,3413	(S) TACTTTG	-3,0868
(E) ACCAATAC	2,4035	(E) GACGAAA	1,8236	(E) AACCATAG	2,2476	(S) CACGGGT	-2,2012	(S) TTTAGGGA	-2,4517	(S) TCTTCGA	-2,6805
(E) ACCATCCC	2,4441	(E) GAGACTA	1,8211	(E) AACCATCA	2,7581	(S) CACTGCG	-4,0919	(S) TTTTCTAG	-2,7624	(S) TGATGTT	-2,5037
(E) ACCATTAC	2,0831	(E) GCGCGAA	1,7163	(E) AACCTTCC	2,8182	(S) CACGGGG	-2,4075	(S) AACGGGGG	-2,8040	(S) TGGACTC	-2,8517
(E) ACCATTTC	2,1483	(E) GCGTACC	2,2399	(E) AACTAAGA	2,2984	(S) CAGTGGC	-2,1777	(S) AAGAGGGG	-3,2438	(S) TGGGGGT	-2,4557
(E) ACCCAAGA	2,2054	(E) GGATATC	1,2158	(E) AAGACTTA	2,6493	(S) CATCGGG	-3,2698	(S) AAGCGGGG	-3,5597	(S) TGGGGTT	-2,3376
(E) ACCCCCCC	2,3777	(E) GGATGAA	1,7618	(E) AATCATCC	3,2004	(S) CATCGGT	-2,4067	(S) AAGGGCGA	-2,6237	(S) TGGTAGG	-2,7076
(E) ACCCCGAC	2,1767	(E) GGCCCAA	1,6723	(E) AATCCATC	2,4363	(S) CATCTTT	-3,1473	(S) AAGGGGGA	-2,5151	(S) TGGGGGG	-2,7020
(E) ACCGATAC	2,0599	(E) GTCCCAT	1,6653	(E) ACAAGAGA	2,2997	(S) CATGGGT	-3,6668	(S) AAGGGGGC	-3,5904	(S) TTTCTTA	-2,7746
(E) ACCGATTA	2,6360	(E) TAAAGAA	2,1291	(E) ACAATCCC	2,4552	(S) CATTTGT	-2,2305	(S) AAGGGGGG	-3,5839	(S) TTTGTTG	-3,1097
(E) ACCGCACA	2,1792	(E) TAAAGAT	2,2247	(E) ACACACAA	2,5278	(S) CATTTGT	-2,2683	(S) ACACGGGA	-2,6167	(S) TTTTCAG	-2,9416
(E) ACCGCATA	2,2871	(E) TAAAGAC	1,8322	(E) AGACAAAC	2,8417	(S) CATTTTG	-2,2090	(S) ACGGGGGC	-2,8745	(S) AGTGTCAT	-2,4829
(E) ACCGCCAA	2,3625	(E) TAAATA	2,0257	(E) AGATCAAA	2,6372	(S) CCAAGT	-2,3535	(S) ACTGCGCG	-2,8570	(S) GAAGGGC	-3,8462
(E) ACCGGAAG	2,1963	(E) TAAAAAT	2,7695	(E) AGTTAATA	2,2528	(S) CCCCCTT	-2,1399	(S) ACTGGGGG	-4,0497	(S) GCGGGCA	-2,8047
(E) ACCGGAAG	2,1963	(E) TAAACAA	1,7651	(E) CAACACAA	2,9508	(S) CCCCCTG	-2,1933	(S) AGAAGTTC	-3,0062	(S) GGGGTAT	-5,1050
(E) ACCTAAGA	2,6691	(E) TAAACCA	1,9648	(E) CAATGAAA	3,0927	(S) CCTAGGG	-2,3800	(S) AGAGGGGA	-2,9109	(S) ACAGTGT	-2,5524
(E) ACCTAATA	3,0812	(E) TAAAGAA	1,7096	(E) CACTAATA	2,5590	(S) CCTAGTG	-2,4999	(S) AGAGGGGG	-3,6969	(S) CATTTGG	-5,0230
(E) ACCTATAA	2,5277	(E) TAAATAC	1,6956	(E) CATAACAA	3,2244	(S) CCTATGG	-2,2085	(S) AGCAAGGG	-2,6891	(S) CCCCCGG	-3,2511

(E)	ACCTCAGC	2,1588	(E)	TAACAAA	1,6503	(E)	CATCAAA	2,4138	(S)	CCTCGTT	-2,2068	(S)	AGCCCGGG	-2,7808	(S)	CGTGGTA	-4,0134
(E)	ACCTCCAC	2,4664	(E)	TAACCTCT	2,0042	(E)	CCCATTAA	2,2847	(S)	CCTGGGG	-2,3351	(S)	AGCCGGGC	-2,6040	(S)	GAGTGCC	-2,8142
(E)	ACGAAAAC	2,6206	(E)	TAAGAAA	1,8276	(E)	GAAATATC	2,3456	(S)	CCTGTCG	-2,1629	(S)	AGCGGGGG	-3,7403	(S)	GAGTTTT	-2,7964
(E)	ACGAATAC	2,9528	(E)	TAAGAAG	2,2238	(E)	GACTACCA	2,2322	(S)	CCTTCGG	-2,8455	(S)	AGCTGGGG	-3,2103	(S)	GCGGGAC	-2,3362
(E)	ACGAATAG	2,0570	(E)	TAGACA	2,5727	(E)	GCAACCAG	2,2769	(S)	CCTTCGT	-2,5706	(S)	AGGAGGGC	-3,7845	(S)	GCTCTTG	-2,3572
(E)	ACGACAAG	2,5461	(E)	TAAGATT	1,9480	(E)	GCCAAAAC	2,2439	(S)	CGCTCAG	-2,3959	(S)	AGGGCCGG	-3,9133	(S)	GCTGACG	-2,7510
(E)	ACGACACC	2,2542	(E)	TAATAAA	2,3084	(E)	TAGACCTA	2,6495	(S)	CGGCCTT	-3,1429	(S)	AGGGGCAA	-3,2259	(S)	GCTTAGT	-2,5905
(E)	ACGACTAC	2,1251	(E)	TAATCTA	1,6628	(E)	TAGCAATA	2,5788	(S)	CGTTGCG	-2,2359	(S)	AGGGGCGA	-3,3981	(S)	GGACAGG	-2,8372
(E)	ACGAGACG	2,1183	(E)	TACAACC	1,9614	(E)	TAGCCAAA	3,0432	(S)	CGTTGGG	-3,0308	(S)	AGGGGCGG	-4,3998	(S)	GGTATTG	-3,6572
(E)	ACGAGTTA	2,4849	(E)	TACAAGA	1,8442	(E)	TATGACCC	2,8706	(S)	CGTTTTG	-2,2153	(S)	AGGGGGAC	-3,7431	(S)	GTACATT	-2,6019
(E)	ACGATACC	3,0247	(E)	TACCCAA	2,4092	(E)	TCCGGAAA	2,4032	(S)	CTGGGGG	-2,6686	(S)	AGGGGGGA	-3,2248	(S)	GTCCAGG	-2,8334
(E)	ACGATTCA	2,5740	(E)	TACCCAT	1,6295	(E)	TCGAGATA	3,7975	(S)	CTGTGCC	-2,5796	(S)	AGGGGGGC	-2,7667	(S)	GTGCAGC	-2,4774
(E)	ACGCCAAG	2,3625	(E)	TACCCCC	1,9463	(E)	TCGCTATA	2,8319	(S)	CTTGTC	-2,1602	(S)	AGGGGGGG	-3,9687	(S)	TCAGGGA	-2,3568
(E)	ACGCAGAA	2,2940	(E)	TACCGAA	2,2975	(E)	TCGTACTG	3,3376	(S)	CTTTAGT	-3,6101	(S)	AGGGGGTG	-2,6240	(S)	TGACAGG	-2,9852
(E)	ACGCCAAC	2,2718	(E)	TACGAAT	1,8702	(E)	AAAAAAGG	2,2534	(S)	CTTTGTC	-2,8822	(S)	AGGGGTCC	-2,8836	(S)	TGAGATT	-2,2783
(E)	ACGCCCCA	2,2548	(E)	TACGATA	2,3363	(E)	AAACAAAA	2,6868	(S)	GAAGGGG	-2,1337	(S)	AGTAGGGG	-3,8670	(S)	TGGACGC	-2,7855
(E)	ACGCTCAA	2,6069	(E)	TACGATC	2,5801	(E)	AAACAGAC	2,7159	(S)	GACCGTT	-2,4204	(S)	AGTGGGAG	-2,4575	(S)	TGTGACG	-2,3394
(E)	ACGGAACA	2,1896	(E)	TACGCAG	2,4179	(E)	AAAACCTG	2,6805	(S)	GACGCTC	-2,4128	(S)	ATCAGGGG	-3,2882	(S)	GGGTCCA	-2,5537
(E)	ACTAACAA	2,2136	(E)	TACGCCG	2,0343	(E)	AAAAGGAA	2,7638	(S)	GAGAGGG	-2,4739	(S)	ATGGCGGA	-3,1362	(S)	GGTATCT	-2,7620
(E)	ACTAACAG	2,6264	(E)	TACGGCA	2,3595	(E)	AAACACAG	2,2883	(S)	GAGCAAG	-2,2180	(S)	ATGGCGGG	-2,7520			
(E)	ACTAACCG	3,2383	(E)	TACGTAT	2,3119	(E)	AAACCGCA	2,7838	(S)	GAGCGGG	-4,2970	(S)	ATGGGGCG	-3,5228			
(E)	ACTACTAA	2,2386	(E)	TACTAAG	1,9142	(E)	AAACCTAC	2,8711	(S)	GAGCGGT	-2,2324	(S)	ATGGGGGA	-3,4623			
(E)	ACTAGAGA	2,1938	(E)	TACTAAT	1,8864	(E)	AAACGCCG	2,4195	(S)	GAGGGCG	-2,8394	(S)	ATGGGGGG	-3,8732			
(E)	ACTAGCCC	2,5139	(E)	TACTAAC	1,7610	(E)	AAAGAAGA	2,4568	(S)	GAGGGTT	-3,0631	(S)	ATGTCCGG	-2,4685			
(E)	ACTAGCTA	2,1630	(E)	TACTAGT	1,7804	(E)	AAAGACCG	3,3717	(S)	GATAAAA	-3,1653	(S)	ATTGGGGA	-3,6978			
(E)	ACTATCAA	2,5063	(E)	TACTATA	2,8187	(E)	AAAGCCAA	2,4891	(S)	GATATCG	-2,3745	(S)	CAAGGGGG	-2,4735			
(E)	ACTCAACA	2,2019	(E)	TACTATC	1,6107	(E)	AAAGCCGA	2,9015	(S)	GATGGGG	-2,2587	(S)	CACAGTTG	-2,4382			
(E)	ACTCATAC	2,5320	(E)	TACTCAA	2,3849	(E)	AAATAATC	2,8139	(S)	GATGTGG	-2,2388	(S)	CACGGGGG	-3,5817			
(E)	ACTCCGAC	2,1206	(E)	TACTGAT	1,7858	(E)	AACAATAG	2,8110	(S)	GATTCTG	-2,1627	(S)	CACGTGGA	-3,6189			
(E)	ACTCTACC	2,2254	(E)	TAGACAT	1,6581	(E)	AACACAAC	3,0792	(S)	GCACGTG	-2,3681	(S)	CAGGGGAG	-3,3738			
(E)	ACTCTACG	2,0922	(E)	TAGATAT	1,7790	(E)	AACCACTA	2,4775	(S)	GCATCTG	-3,0509	(S)	CAGGGGGC	-4,2707			
(E)	ACTGAATC	2,0576	(E)	TAGATCC	2,0404	(E)	AACCAGAG	2,3181	(S)	GCAGGGG	-2,9629	(S)	CATGGGGG	-3,1963			
(E)	ACTGACCC	2,8382	(E)	TAGATTG	1,7902	(E)	AACGACTG	2,2491	(S)	GCAGGGT	-2,2394	(S)	CCAGGGTG	-3,2629			
(E)	ACTGACGA	2,3026	(E)	TAGCAAA	2,0301	(E)	AACGAGAA	2,6112	(S)	GCAGGTG	-2,2212	(S)	CCCGGGGC	-4,0495			
(E)	ACTGACGC	2,1962	(E)	TAGCCAC	1,8383	(E)	AACGATAA	2,6634	(S)	GCATGGT	-3,0130	(S)	CCCGGGGG	-2,6578			
(E)	ACTTAATA	2,4157	(E)	TATCCAA	1,9337	(E)	AACGCCAA	2,6776	(S)	GCCTTTT	-2,4325	(S)	CCGGGATC	-2,6293			
(E)	ACTTAGTG	2,2755	(E)	TATCGAC	1,6658	(E)	AACATATC	2,6028	(S)	GCCGGGG	-2,1745	(S)	CCTAGGGA	-2,8088			
(E)	ACTTCCCA	2,4980	(E)	TATTAAA	2,2200	(E)	AACCTCAA	3,0803	(S)	GCGACGG	-2,1838	(S)	CCTGCGGG	-2,8747			
(E)	AGAAAACA	3,5056	(E)	TATTATA	2,3840	(E)	AACCTGAC	2,5414	(S)	GCGAGAG	-3,2642	(S)	CCTGGGGG	-4,0414			
(E)	AGAAAACG	2,5273	(E)	TATTTTT	1,7467	(E)	AAGAACGA	3,2632	(S)	GCGATCG	-2,5265	(S)	CGAAGGGA	-3,1704			
(E)	AGAAAACC	3,3292	(E)	TCAAAT	2,7562	(E)	AATAAAC	3,6122	(S)	GCGCACT	-2,6216	(S)	CGAAGGGC	-3,0893			
(E)	AGAAAACG	2,0983	(E)	TCACATA	2,0856	(E)	AATAATTG	2,4458	(S)	GCGCATG	-2,1575	(S)	CGACGGGG	-4,3481			
(E)	AGAAAACG	2,2059	(E)	TACACGA	2,4987	(E)	AATACCAA	2,8380	(S)	GCGCGTG	-3,6741	(S)	CGAGGGAG	-3,1422			
(E)	AGAACCAC	2,6192	(E)	TCAGACT	1,7225	(E)	AATACGGA	2,4959	(S)	GCGGAGG	-2,9877	(S)	CGATGGCG	-2,4432			
(E)	AGAAGCCA	2,0920	(E)	TCAGATT	1,9038	(E)	AATAGATC	2,3069	(S)	GCGGCAT	-2,3396	(S)	CGCCATGC	-2,8833			
(E)	AGACAAAA	2,3393	(E)	TCAGCAA	1,8371	(E)	AATCACAA	3,5408	(S)	GCGGGCC	-2,5433	(S)	CGCCCCGG	-3,5001			
(E)	AGACAAAC	2,8337	(E)	TCAGCGC	2,3132	(E)	AATCCAAC	3,5168	(S)	GCGTGTG	-2,1384	(S)	CGCCCGTG	-2,7673			
(E)	AGACATAA	2,6565	(E)	TCAGCGG	2,4592	(E)	AATCCATG	2,3198	(S)	GCGTGTG	-2,1627	(S)	CGCCGTGG	-2,6945			
(E)	AGACATCA	2,4199	(E)	TCAGTCC	2,2096	(E)	AATGCAAA	2,5132	(S)	GCTAAC	-2,2444	(S)	CGCTCTTG	-2,7076			
(E)	AGACCAAG	2,8686	(E)	TCATAAT	1,9446	(E)	AATTACCC	3,2852	(S)	GCTATTG	-2,3345	(S)	CGCGGGGG	-4,8442			
(E)	AGACTATA	2,4335	(E)	TCCAAAG	2,8554	(E)	ACAAATC	2,2803	(S)	GCTCAGC	-2,4236	(S)	CGCGGGGA	-3,3934			
(E)	AGAGAAAC	2,2732	(E)	TCCACAA	2,2932	(E)	ACAAATCG	2,2786	(S)	GCTCAGG	-2,1249	(S)	CGCGGGGG	-2,8385			
(E)	AGAGAATC	3,3426	(E)	TCCCAAT	2,5194	(E)	ACAAAGAC	2,7406	(S)	GCTCGGG	-2,3969	(S)	CGCGTCTG	-3,5447			
(E)	AGAGCACA	2,1576	(E)	TCCCAAC	2,1908	(E)	ACAATACC	2,6328	(S)	GCTCGCT	-2,1504	(S)	CGCGTGTG	-2,7350			
(E)	AGAGCACC	2,4716	(E)	TCCGACT	1,8146	(E)	ACAATATC	2,7679	(S)	GCTCTTA	-2,2930	(S)	CGCGTTTG	-2,6635			
(E)	AGATAAAC	2,1342	(E)	TCCGAGA	1,8661	(E)	ACACCACC	2,3658	(S)	GCTGCGC	-2,1549	(S)	CGCTGGGG	-4,0266			
(E)	AGATCAAA	2,8012	(E)	TCCGAGC	2,4931	(E)	ACACCATA	2,3119	(S)	GCTGTTG	-2,3244	(S)	CGCTGTGG	-2,5556			
(E)	AGATCATA	2,1744	(E)	TCCGATA	1,7764	(E)	ACAGTTTA	3,0590	(S)	GCTTCGC	-2,3681	(S)	CGCTTCGG	-2,8027			
(E)	AGATCTTA	2,5299	(E)	TCCGCAA	2,2341	(E)	ACATAATC	2,2582	(S)	GCTTCGG	-2,9469	(S)	CGCTTGGG	-2,5581			
(E)	AGATTGGA	2,6116	(E)	TCCGGAC	2,4993	(E)	ACCCACAA	2,9805	(S)	GCTTGCG	-2,1778	(S)	CGGAGGGG	-3,7712			
(E)	AGATTTAA	2,5828	(E)	TCCGGAG	1,8882	(E)	ACCCAAAA	2,9484	(S)	GCTTTTG	-2,1437	(S)	CGGGGGAT	-3,4955			
(E)	AGATTTC	2,2054	(E)	TCCTAAA	1,8863	(E)	ACCTCACC	2,9680	(S)	GGACATG	-2,1765	(S)	CGGGGCGC	-2,9087			
(E)	AGCAAAAA	3,6978	(E)	TCCTACA	2,0913	(E)	ACGAAACG	2,4092	(S)	GGAGAGG	-2,2623	(S)	CGGGGCGG	-2,7731			
(E)	AGCAAAAC	2,0941	(E)	TCCTATC	1,7750	(E)	ACGAAGCC	2,5453	(S)	GGAGCCG	-2,4900	(S)	CGGGGGCA	-3,2599			
(E)	AGCCCGAA	2,0780	(E)	TCGAAAT	1,9702	(E)	ACGAGAAA	2,4604	(S)	GGAGCTG	-2,7016	(S)	CGGGGGGA	-2,4499			
(E)	AGCCTCAA	2,4106	(E)	TCGAAGC	2,3706	(E)	ACGAGGAA	2,4483	(S)	GGAGGTT	-2,2682	(S)	CGGGGGGC	-3,9234			
(E)	AGCGAACG	2,0820	(E)	TCGCAAT	3,0154	(E)	ACGGCAAC	2,4012	(S)	GGATACG	-2,2960	(S)	CGGGGGGG	-3,7417			
(E)	AGCGCTCA	2,1197	(E)	TCGCAAT	2,4007	(E)	ACGTTCCC	2,4791	(S)	GGATCAC	-2,2419	(S)	CGGGGGTG	-3,4325			
(E)	AGCTACCA	2,4463	(E)	TCGCAAG	2,1144	(E)	ACTAACGA	2,3214	(S)	GGATCGC	-2,2064	(S)	CGGGGTGA	-3,6638			
(E)	AGGAACCC	3,0820	(E)	TCGCATA	1,8418	(E)	ACTACCCC	2,8110	(S)	GGATCGG	-2,2955	(S)	CGGGGTTA	-4,2226			
(E)	AGTAACAC	2,4587	(E)	TCGCCAG	1,6174	(E)	ACTTACTA	2,5125	(S)	GGCAACG	-2,2240	(S)	CGGTGGGG	-4,2364			
(E)	AGTAAGAC	2,5679	(E)	TCGCCAT	3,1082	(E)	AGAACCAAC	3,1775	(S)	GGCACGG	-2,4053	(S)	CGGTGTGG	-2,7092			
(E)	AGTAACCC	2,0913	(E)	TCGCCCA	2,0357	(E)	AGAACACA	2,8718	(S)	GGCACCT	-2,9363	(S)	CGGTTTGG	-2,4558			
(E)	AGTAGAAC	2,9146	(E)	TCGCCGA	1,8553	(E)	AGAAGAAA	2,4829	(S)	GGCAGTC	-2,5901	(S)	CGTACGTC	-3,4662			
(E)	AGTCCAAA	2,2531	(E)	TCGCTCC	2,4864	(E)	AGATAACA	4,2117	(S)	GGCATGT	-2,3677	(S)	CGTATGCC	-2,9160			
(E)	AGTCCAC	2,2085	(E)	TCGCTGG	1,6252	(E)	AGATAAGA	2,6372	(S)	GGCATTC	-2,3819	(S)	CGTCACTG	-2,7558			
(E)	AGTGAATC	2,0659	(E)	TCGGAAA	2,6875	(E)	AGATAGAC	2,5294	(S)	GGCCATG	-3,2914	(S)	CGTCGAGA	-2,6518			
(E)	AGTGACAC	2,1245	(E)	TCGGAGT	1,6798	(E)	AGATATAC	2,3576	(S)	GGCCCAAC	-2,1300	(S)	CGTCGTGC	-2,5888			
(E)	AGTTAATA	2,2953	(E)	TCGGCCC	1,6171	(E)	AGATCCAA	2,4271	(S)	GGCCTGT	-2,5211	(S)	CGTGCTTA	-2,7247			
(E)	AGTTACAA	3,3719	(E)	TCGTACC	1,6331	(E)	AGATTCTC	2,7516	(S)	GGCTTTG	-3,5222	(S)	CGTGGGAA	-3,6083			
(E)	ATACAAAA	3,5136	(E)	TCGTAGA	1,7878	(E)	ATCACAAAC	3,2041	(S)	GGGACCG	-2,4010	(S)	CGTGGGCA	-3,7904			



(E) ATACAAAC	2,5201	(E) TCGTCGT	2,5716	(E) ATCACATA	2,2863	(S) GGGACGA	-2,8856	(S) CGTGGGCG	-4,0030		
(E) ATACCACA	2,0683	(E) TCTCGAG	1,7516	(E) ATCTAACA	2,6410	(S) GGGAGGA	-2,8174	(S) CGTGGGGA	-2,5223		
(E) ATACGACG	2,4963	(E) TGA AAC	1,9504	(E) ATGCAAAA	2,3651	(S) GGGAGGC	-2,3009	(S) CGTGGGGG	-3,1006		
(E) ATACTCAC	2,4737	(E) TGAATA	1,7266	(E) ATTACTAC	2,6876	(S) GGGCACT	-2,4148	(S) CGTTGTGG	-3,1100		
(E) ATACTCCA	2,5704	(E) TGAATCA	1,9752	(E) CAAAATTA	3,3209	(S) GGGCATC	-2,8727	(S) CTACGGGG	-3,4179		
(E) ATATAAGA	2,8423	(E) TGACACA	1,6380	(E) CAAACCTG	2,2439	(S) GGGCCAA	-2,6157	(S) CTACGGTG	-2,9849		
(E) ATATACCA	2,6785	(E) TGACACT	1,8768	(E) CAAGAAAC	2,5027	(S) GGGCCCG	-2,7718	(S) CTCAGTGC	-3,5292		
(E) ATATCAAC	2,2964	(E) TGACCGA	1,8688	(E) CAATACAA	2,5516	(S) GGGCCGG	-2,9724	(S) CTCGGGGG	-4,4366		
(E) ATATCACC	2,1799	(E) TGACTAC	2,9532	(E) CACCAATC	2,9107	(S) GGGCCGT	-3,4159	(S) CTCGGGGG	-3,4116		
(E) ATATCCGA	2,3210	(E) TGACCAA	1,9643	(E) CAGAACAA	2,3833	(S) GGGCGAG	-2,1961	(S) CTCGGTGC	-3,2731		
(E) ATATTTAA	2,3834	(E) TGATCAA	1,8621	(E) CAGACACA	2,2567	(S) GGGCCGC	-2,2509	(S) CTCGTGGG	-2,9115		
(E) ATCAAACA	3,6911	(E) TGATGAA	1,8301	(E) CAGATTAC	2,3554	(S) GGGCGGG	-4,4848	(S) CTCCTGGC	-2,5070		
(E) ATCAATAA	3,7426	(E) TGATTTA	1,8382	(E) CATCACCC	2,2837	(S) GGGCTGG	-2,5297	(S) CTCCTGGT	-2,7327		
(E) ATCCAATA	2,2148	(E) TGGCGAA	2,8221	(E) CATTGATA	2,3972	(S) GGGGAGA	-2,3954	(S) CTCCTGTC	-2,5625		
(E) ATCCACCC	2,3435	(E) TGGCGAG	2,0933	(E) CAAAAAGA	2,4995	(S) GGGGAGC	-2,1885	(S) CTGGCCGG	-4,4937		
(E) ATCCATAA	2,5308	(E) TGGCGAA	2,0068	(E) CCAAGGCC	2,2415	(S) GGGGATT	-2,2727	(S) CTGGGGCC	-3,1551		
(E) ATCCGAAA	2,0733	(E) TGGAAAC	1,7232	(E) CCACAACA	2,5494	(S) GGGGCCA	-4,4318	(S) CTGGGGAA	-3,3082		
(E) ATCCTCAG	2,0575	(E) TGGCGCC	1,6255	(E) CCCCCCAA	2,9394	(S) GGGGCCG	-2,7553	(S) CTGGGGAG	-2,9843		
(E) ATCGACTA	2,1603	(E) TTACCCG	2,0385	(E) CCCCTCA	2,3488	(S) GGGGCGT	-2,9786	(S) CTGGGGCG	-3,0111		
(E) ATCGTCCC	2,6514	(E) TTATACA	1,9848	(E) CCCGAAGG	2,9473	(S) GGGGCTA	-3,4020	(S) CTGGGGGC	-4,3029		
(E) ATCTACAC	2,4870	(E) TTATGAA	1,6987	(E) CCTACAA	2,2333	(S) GGGGCTT	-2,6438	(S) CTGGGGGG	-2,7024		
(E) ATCTAGAA	2,8010	(E) TTCACCC	1,9370	(E) CCGACACG	2,4868	(S) GGGGGAT	-5,1977	(S) CTGTGCGC	-3,3378		
(E) ATCTCGAA	2,0624	(E) TTCCAAA	1,6433	(E) CCGCAGAA	2,4105	(S) GGGTAGG	-2,4783	(S) CTGTGGGC	-3,0965		
(E) ATGCAATA	2,3140	(E) TTCCAAG	1,6761	(E) CCGCCAAA	2,2854	(S) GGGTAGT	-2,7916	(S) CTGTGGGG	-2,6902		
(E) ATGTCATC	2,0741	(E) TTCCATG	1,8152	(E) CGAAAAAA	2,4746	(S) GGGTCTA	-2,1989	(S) CTTCTGGT	-2,6173		
(E) ATTAAGCC	2,1121	(E) TTCGAAC	2,3661	(E) CGAAAAAC	2,2833	(S) GGGTCTC	-2,3505	(S) CTTCTGGG	-2,7853		
(E) ATTAAGAC	2,2319	(E) TTCGACA	2,0817	(E) CGAAATAA	2,2436	(S) GGGTGCC	-2,3700	(S) CTTCTGTT	-2,5950		
(E) ATTAAGAG	2,5036	(E) TTCGAGT	1,7861	(E) CGAACAAA	2,2363	(S) GGGTGGA	-2,7439	(S) CTTCTGTT	-3,0636		
(E) ATTACATA	2,2114	(E) TTCGCAG	2,4671	(E) CGAAGACA	2,6361	(S) GGGTGGT	-2,6451	(S) CTTGGGGA	-2,4728		
(E) ATTACCCC	2,5899	(E) TTCGCCC	2,0325	(E) CGAGACTA	2,2928	(S) GGGTGGT	-2,6779	(S) GAAGGGGA	-2,5676		
(E) ATTACGAC	2,2269	(E) TTCGGCC	1,7541	(E) CGATATAC	2,3493	(S) GGGTTGC	-2,1536	(S) GAAGGGGG	-3,0935		
(E) ATTACTGA	2,3692	(E) TTCGTCA	1,8184	(E) GAAATTTG	2,7615	(S) GGGTTGG	-2,6306	(S) GACCGGGG	-3,2277		
(E) ATTACTTA	2,3452	(E) TTCTAGC	1,7081	(E) GAAATTTAG	2,4002	(S) GGTATCG	-3,1901	(S) GAGCGGGT	-2,6710		
(E) ATTAGATC	2,0916	(E) TTCTCAA	1,7388	(E) GAAATTTCC	2,4565	(S) GGTCAAC	-2,1757	(S) GAGGGCAG	-2,8393		
(E) ATTAGCCA	2,4400	(E) TTGCCAA	1,8787	(E) GAACACCA	3,5135	(S) GGTACAG	-2,6012	(S) GAGGGGAA	-3,0110		
(E) ATTATCTA	2,1437	(E) TTGCCCC	1,7586	(E) GAACACCG	2,3296	(S) GGTACAG	-2,3619	(S) GAGGGGAC	-3,5707		
(E) ATTCACCA	2,5711	(E) TTGCCCT	1,7743	(E) GAACACACA	3,1737	(S) GGTACATA	-2,1494	(S) GAGGGGCG	-2,8439		
(E) ATTCCTAC	2,4885	(E) TTGCCGA	2,3459	(E) GAACCCGA	2,2990	(S) GGTCTAT	-2,1370	(S) GAGGGGGA	-3,3871		
(E) ATTCCTCC	2,4333	(E) TTGCTCC	1,9581	(E) GAACCTCA	2,3059	(S) GGTCTCA	-2,5501	(S) GAGGGGGG	-3,7371		
(E) ATTTAGGA	2,4336	(E) TTGGATC	1,7078	(E) GAAGTGCA	2,3193	(S) GGTCTCG	-2,7942	(S) GAGTGGGG	-3,2267		
(E) ATTTATAA	2,0668	(E) TTGGCCC	1,9620	(E) GAAGTTAA	2,5210	(S) GGTGAAG	-2,7573	(S) GATCGGGG	-2,7220		
(E) ATTTTACA	2,0676	(E) TTGTGAC	1,7352	(E) GAATAATA	2,3842	(S) GGTGCAA	-2,3397	(S) GATGAGGG	-2,6862		
(E) CAAAAAAG	2,1090	(E) TTGTGCA	1,6960	(E) GAATTATA	2,6019	(S) GGTGCGA	-2,1798	(S) GATGGGAA	-2,7205		
(E) CAAAATC	2,3936	(E) TTTAATC	2,3412	(E) GACAAATC	2,9649	(S) GGTGGAA	-2,9321	(S) GATGGGGA	-2,9013		
(E) CAAACAAA	2,4145	(E) TTATATC	1,6276	(E) GACGAGCC	2,2757	(S) GGTGGAC	-2,5947	(S) GATGGGGG	-4,2287		
(E) CAAACACC	2,4304	(E) TTTCAC	2,0510	(E) GATTACAC	2,5451	(S) GGTGTGC	-3,0839	(S) GATGGTGC	-2,9559		
(E) CAAACTAA	2,8410	(E) TTTCGCC	1,9206	(E) GATTGAGA	2,2778	(S) GGTGTTT	-3,2327	(S) GATTGGGG	-3,5741		
(E) CAAACTAC	2,9654	(E) TTTGACT	1,7105	(E) GCAAAAAC	3,7797	(S) GGTATTG	-2,1222	(S) GCAGGGGG	-2,5056		
(E) CAAGAAAG	2,2643	(E) TTGTGAT	2,1338	(E) GCAACCCA	2,3042	(S) GGTTCAT	-3,4271	(S) GCATGGGG	-2,6040		
(E) CAAGAGAC	2,1718	(E) TTGCGG	2,1091	(E) GCCCAAAC	2,6819	(S) GGTTCCT	-3,1160	(S) GCCCGGGG	-2,5721		
(E) CAAGAGCC	2,3853	(E) TTTCGCG	2,4457	(E) GCCTATAA	2,2431	(S) GGTTCGT	-2,1230	(S) GCCCGGGG	-3,8757		
(E) CAAGATAA	2,6505	(E) TTTTACC	2,6874	(E) GCCTCCCC	2,4290	(S) GGTTCCT	-2,2934	(S) GCGGGGGA	-3,3797		
(E) CAAATATA	2,3706	(E) TTTTAGA	1,6289	(E) GCGAAAAG	2,5590	(S) GTACATC	-3,1582	(S) GCGGGGGG	-3,7493		
(E) CAAATATA	3,4342	(E) AAACAGT	-1,6603	(E) GCGAAATA	3,4372	(S) GTACCGG	-2,6371	(S) GCTCGGAC	-2,4497		
(E) CAACAATA	2,3692	(S) AACGGTT	-1,8898	(E) GCGAAATC	2,8056	(S) GTACGGG	-2,5108	(S) GCTCGGGA	-2,6525		
(E) CAACACCG	2,4098	(S) AAGCGGT	-2,3224	(E) GCGAATGC	2,4771	(S) GTACGTG	-2,3474	(S) GCTGGGGG	-3,1671		
(E) CAACATCC	2,2441	(S) AAGGCC	-1,7272	(E) GCGCCAAA	2,6346	(S) GTACTCC	-2,3782	(S) GGAAGGGA	-3,0470		
(E) CAACCGAA	2,2372	(S) AAGCGGT	-1,6759	(E) GCGCCCCA	2,4154	(S) GTATCGT	-2,3667	(S) GGAAGGGG	-4,0437		
(E) CAACCTCA	2,1736	(S) AAGGGAC	-2,1018	(E) GCGCCGGA	2,2837	(S) GTCAAGG	-2,6082	(S) GGACGCGG	-2,7879		
(E) CAAGAAAG	2,1554	(S) AAGGGCA	-2,3762	(E) GCGCTGCA	3,2439	(S) GTCCGGG	-2,3387	(S) GGAGGGAA	-3,2513		
(E) CAAGAAATC	2,1881	(S) AAGGGGC	-3,6032	(E) GCGCTGTA	2,3812	(S) GTCCGTC	-2,1699	(S) GGAGGGAG	-3,2429		
(E) CAAGACCC	2,8736	(S) AAGGGGT	-2,5159	(E) GTATAATC	2,2412	(S) GTCCGTT	-2,3267	(S) GGAGGGCC	-3,2894		
(E) CAAGCATA	2,1030	(S) AAGGTGG	-2,9607	(E) GTCTATAA	2,6088	(S) GTCTATA	-2,1398	(S) GGAGGGGC	-4,0805		
(E) CAAGCCCA	2,8705	(S) AAGTAGG	-1,6129	(E) TAAAAAGA	2,2343	(S) GTCTCAC	-2,3945	(S) GGATCCGG	-2,5646		
(E) CAATAAAA	3,1161	(S) AATCGGT	-2,1481	(E) TAAACCC	3,1661	(S) GTCTTCC	-2,1714	(S) GGATGGCG	-3,6933		
(E) CAATAACA	2,6392	(S) AATGGTC	-2,0623	(E) TAAACCCC	2,6254	(S) GTCTTCG	-2,5244	(S) GGCACCGG	-3,0039		
(E) CAATAGAC	2,9041	(S) ACAAGGT	-2,4098	(E) TAAACTTC	2,4981	(S) GTCTTGC	-2,2277	(S) GGCACGGG	-2,8322		
(E) CAATATCC	2,4335	(S) ACACAGT	-1,8594	(E) TAACCCAC	2,2731	(S) GTCTTGT	-2,1958	(S) GGCAGGGA	-3,8849		
(E) CAATCCAG	2,4425	(S) ACAGGCG	-2,0155	(E) TAAGCAAA	2,7555	(S) GTGCAAC	-2,4911	(S) GGCAGGGC	-3,9712		
(E) CAATGAAA	2,5727	(S) ACAGGTT	-2,1488	(E) TAATAAAC	2,4128	(S) GTGCAGA	-2,4593	(S) GGCCCGCT	-2,7047		
(E) CAATGAAC	2,3543	(S) ACCAGGT	-2,1253	(E) TACGAACC	2,4083	(S) GTGCAGT	-2,6703	(S) GGCCGGGC	-3,8014		
(E) CAATTCCC	2,3268	(S) ACCCGTA	-1,7757	(E) TACGACCA	4,2230	(S) GTGCATT	-2,2221	(S) GGCCGGGG	-3,3408		
(E) CAATTTGA	2,3101	(S) ACCTGTG	-1,6210	(E) TACGCAAA	2,2538	(S) GTGCCCA	-2,5538	(S) GGCGCCTG	-3,0049		
(E) CACAAATC	2,9320	(S) ACGACTG	-2,0796	(E) TAGACAC	2,3762	(S) GTGCCCG	-2,5707	(S) GGCGGGGC	-3,6885		
(E) CACAACCA	2,3423	(S) ACGACTT	-2,0468	(E) TAGAGAGA	2,6176	(S) GTGCCAG	-2,2049	(S) GGCGGGGG	-3,2668		
(E) CACAACCC	3,1122	(S) ACGAGGT	-2,2559	(E) TAGAGCTC	2,5132	(S) GTGCCGC	-4,4871	(S) GGCGGTGG	-2,9800		
(E) CACACCTC	2,4549	(S) ACGATCA	-1,7084	(E) TAGCAGCG	2,3057	(S) GTGCTCT	-4,0604	(S) GGCGTCTC	-3,4527		
(E) CACATACC	2,4584	(S) ACGATCC	-1,6470	(E) TAGCCAAG	3,5649	(S) GTGCCAG	-2,3388	(S) GGCTGTTC	-2,5589		
(E) CACCACAA	2,5512	(S) ACGCCTA	-2,0230	(E) TATAAGAC	2,5263	(S) GTGGCGA	-2,2930	(S) GGGACGCG	-3,6598		
(E) CACCACCTC	2,1909	(S) ACGGATC	-1,9462	(E) TCAAAAAG	3,5401	(S) GTGGCTC	-2,4693	(S) GGGACGGG	-2,6840		
(E) CACGACAG	2,4720	(S) ACGGGCT	-1,9714	(E) TCAAAAAG	2,3424	(S) GTGGGAT	-2,2108	(S) GGGATTGG	-3,4905		

(E) CACGAGAC	2,7404	(S) ACGGGGG	-1,7603	(E) TCAGCGAG	2,3731	(S) GTGGGGA	-2,3313	(S) GGGCCGGA	-2,6465		
(E) CACGGACG	2,4064	(S) ACGGGTA	-2,2586	(E) TCATATAA	2,2719	(S) GTGGTCG	-2,5291	(S) GGGCCGGG	-2,9367		
(E) CACTAAAA	2,5328	(S) ACGGGTT	-1,8143	(E) TCATCCAC	2,3114	(S) GTGTAGC	-2,2486	(S) GGGCGGGG	-2,6342		
(E) CAGACAGC	2,5369	(S) ACGGTGG	-2,8760	(E) TCCGAGAC	2,7627	(S) GTGTCCT	-3,0372	(S) GGGCGGTG	-3,6014		
(E) CAGACCCC	3,2228	(S) ACGGTTT	-1,8969	(E) TCGAAACG	2,2987	(S) GTGTGCC	-2,4912	(S) GGGGAACG	-3,6449		
(E) CAGACTCC	2,9919	(S) ACGTCGA	-1,8486	(E) TCGACCTG	2,6080	(S) GTGTGGA	-4,2037	(S) GGGGACTC	-2,9358		
(E) CAGAGCCA	2,7662	(S) ACTAACG	-1,7465	(E) TCGATTAC	2,6935	(S) GTGTTTA	-2,5760	(S) GGGGAGGG	-2,7760		
(E) CAGCAAGA	2,2202	(S) ACTAGGT	-1,7402	(E) TCGCACGC	2,3669	(S) GTTAGCT	-2,5095	(S) GGGGATTC	-3,5855		
(E) CAGCACCC	3,3180	(S) ACTATCT	-2,1559	(E) TCGCATCA	2,3658	(S) GTTCACG	-2,8258	(S) GGGGCCAA	-3,1865		
(E) CAGCACTA	2,2860	(S) ACTCAGG	-1,6095	(E) TCGCGAAC	2,9748	(S) GTTCCTG	-2,2889	(S) GGGGCCCG	-3,3900		
(E) CAGGAGCA	2,3556	(S) ACTCGCG	-1,9513	(E) TCGGAATC	2,3889	(S) GTTGAAC	-3,4166	(S) GGGGCCGA	-3,6036		
(E) CAGTCAGA	2,0732	(S) ACTCGGT	-1,9134	(E) TCGGACTG	2,3143	(S) GTTGCCG	-2,4036	(S) GGGGCGCG	-3,8767		
(E) CATAAAGA	2,5992	(S) ACTGAGG	-1,7634	(E) TCGGCCCA	2,6593	(S) GTTGGTA	-3,4329	(S) GGGGCCGA	-3,0368		
(E) CATAACAA	2,8655	(S) ACTGCGC	-1,9654	(E) TCGGCCGA	2,6117	(S) GTTGTGC	-2,8608	(S) GGGGCGCG	-3,7445		
(E) CATACATC	2,9208	(S) ACTGGGC	-1,7821	(E) TCGGCCGG	3,2358	(S) GTTTGAT	-3,1772	(S) GGGGCCGG	-3,0192		
(E) CATAGAGA	2,1686	(S) ACTGTGC	-1,6141	(E) TCGGGAGG	2,2324	(S) TACAGTG	-2,2409	(S) GGGGCTGT	-2,5263		
(E) CATATCAA	2,4758	(S) ACTGTAT	-1,8833	(E) TCGGGGTC	2,6845	(S) TACTCTG	-2,4228	(S) GGGGCTGC	-2,7313		
(E) CATCAAAC	2,5061	(S) ACTGTGC	-2,8472	(E) TCGGTCCA	2,9334	(S) TAGGGTT	-2,9046	(S) GGGGCTTA	-3,7452		
(E) CATCAACA	2,6684	(S) ACTGTTT	-1,6152	(E) TCGGTGGG	3,0241	(S) TAGTCGT	-3,6430	(S) GGGGCTTC	-2,7643		
(E) CATGACAC	2,1861	(S) ACTTCGT	-1,7618	(E) TCGTAGCC	2,5266	(S) TATGGTG	-3,3083	(S) GGGGGACC	-3,7274		
(E) CATTAATA	2,2207	(S) ACTTGGG	-1,9746	(E) TCGTCACG	2,9630	(S) TCCTGTG	-2,2212	(S) GGGGGACG	-3,4778		
(E) CATTCATA	2,3668	(S) AGACGGT	-2,1803	(E) TCGTCCCG	4,2095	(S) TCTCCTC	-2,4803	(S) GGGGGATC	-2,8769		
(E) CATTCACA	2,0879	(S) AGGACTC	-1,8726	(E) TCGTGCGG	2,6916	(S) TCTTCGG	-2,4969	(S) GGGGGCAC	-4,2136		
(E) CCAAAACC	2,2547	(S) AGGCCCT	-2,1350	(E) TCTCCACC	2,3542	(S) TGACGGT	-2,1326	(S) GGGGGCAG	-4,1491		
(E) CCAACTTA	2,1905	(S) AGGCCGG	-1,6798	(E) TCTCCCTG	2,2930	(S) TGACTTA	-2,7346	(S) GGGGGCGC	-2,5836		
(E) CCAAGAAC	2,6044	(S) AGGCTTC	-1,6475	(E) TGAATAA	2,6372	(S) TGCCCCG	-2,4477	(S) GGGGGCGG	-4,3561		
(E) CCAAGACA	2,1828	(S) AGGGACT	-1,9157	(E) TGAATAG	2,2894	(S) TGGGAGG	-2,1970	(S) GGGGGCTA	-3,5685		
(E) CCAAGAGA	2,2427	(S) AGGGGGA	-1,6306	(E) TGACCCCTG	2,3911	(S) TGGGGGC	-2,1377	(S) GGGGGGAC	-4,0826		
(E) CCACAATG	2,1470	(S) AGGGGGT	-2,7052	(E) TGAGCCAA	2,2964	(S) TGTCCGA	-2,2282	(S) GGGGGGGC	-2,5877		
(E) CCACACTG	2,2305	(S) AGGGGTA	-3,3651	(E) TGGGAACC	2,4300	(S) TGTCCGG	-2,2621	(S) GGGGGGGG	-3,7093		
(E) CCACCAAA	2,1350	(S) AGGGTGT	-3,3182	(E) TGGTCCCA	2,4453	(S) TGTGTTT	-2,4390	(S) GGGGGGTC	-2,5750		
(E) CCACCTAC	2,6424	(S) AGGGTGG	-3,0100	(E) TTGCACCA	2,6717	(S) TGTTCCTG	-3,2586	(S) GGGGGGTG	-2,4529		
(E) CCAGATTG	2,1739	(S) AGGTCAT	-2,3514	(E) TTGCCCGC	2,2965	(S) TTCCTCG	-2,2057	(S) GGGGGTCA	-4,2593		
(E) CCAGCCAA	2,3145	(S) AGGTCTC	-1,7901	(E) TTGCCCGG	2,8474	(S) GGTATTT	-2,9352	(S) GGGGGTGA	-4,0325		
(E) CCCAAACC	2,2877	(S) AGGTGTC	-2,2929	(E) TTGCCGCA	2,4748	(S) GGTGGA	-2,2283	(S) GGGGGTGC	-4,0601		
(E) CCCAATCC	2,5331	(S) AGGTGCG	-2,5735	(E) TTGACTA	2,4596	(S) GGTGTGT	-2,5797	(S) GGGGGTGG	-3,3820		
(E) CCCACCAC	2,0673	(S) AGGTCTA	-1,7694	(E) TTGCCAGA	2,6381	(S) TGTGTCA	-2,1550	(S) GGGGGTTC	-4,4159		
(E) CCCATTAA	2,3358	(S) AGGTCTC	-2,6181	(E) TTTGAAAA	3,1750			(S) GGGGTCCG	-3,8779		
(E) CCCCCCGG	2,5198	(S) AGGTCTT	-1,7799	(S) AAGGGGAA	-2,7096			(S) GGGGTTC	-3,7748		
(E) CCCCCGAA	2,4896	(S) AGGTGAG	-1,6606	(S) AGCACTTC	-2,7336			(S) GGGGTTGG	-3,2443		
(E) CCCCCGCC	2,1996	(S) AGGTGAT	-2,5774	(S) AGCCCGTG	-2,7092			(S) GGTATGGG	-2,4908		
(E) CCCCCTCA	2,0821	(S) AGGTGGT	-2,6785	(S) AGGACTTC	-3,1551			(S) GGTCCGAG	-3,1144		
(E) CCCGACTA	2,0652	(S) AGGTGGG	-2,3219	(S) AGGGCGTC	-2,3953			(S) GGTCTGGG	-2,4440		
(E) CCCTAACCA	2,3272	(S) AGGTGTG	-2,2508	(S) AGTTGGGC	-2,8599			(S) GGTGCACG	-3,6155		
(E) CCGACACA	2,7010	(S) AGGTGTT	-2,1494	(S) ATCGTGGG	-2,8227			(S) GGTGGGAG	-3,3163		
(E) CCGACCCA	2,7268	(S) AGGTTC	-2,2284	(S) ATCTGGGC	-2,6487			(S) GGTGGGCA	-3,6444		
(E) CCGGACCC	2,6798	(S) AGGTGTC	-2,4552	(S) ATTCTGGG	-2,2933			(S) GGTGGGGA	-4,1480		
(E) CCTAAAAA	2,2729	(S) AGGTGTC	-2,6076	(S) CAGGTGTC	-2,4422			(S) GGTGTGCG	-2,5744		
(E) CCTAATAC	2,3264	(S) AGTACGC	-1,7200	(S) CATCGGTG	-2,5149			(S) GGTTGCCG	-2,7163		
(E) CCTACCAA	2,9008	(S) AGTAGGC	-2,3172	(S) CCTGGGCG	-2,8829			(S) GTACGGGA	-4,0383		
(E) CCTACTAG	2,3226	(S) AGTCGCG	-2,2243	(S) CCTAATTG	-2,3685			(S) GTACGTCC	-2,4926		
(E) CCTAGATA	2,3287	(S) AGTCTCA	-1,6970	(S) CCTGGCCC	-2,9067			(S) GTACTCGG	-2,4398		
(E) CCTCCACA	2,0842	(S) AGTGAGC	-2,4943	(S) CCTTGCTG	-2,4545			(S) GTATCTTG	-3,0364		
(E) CGAAACAC	2,6383	(S) AGTGAGT	-1,7328	(S) CCTTGCTC	-2,2350			(S) GTATGCGC	-2,8902		
(E) CGAAACCG	2,1397	(S) AGTGGCT	-2,0092	(S) CGCGCGCG	-3,5504			(S) GTCACGGG	-2,4515		
(E) CGAAAGAA	2,4535	(S) AGTGGGT	-1,8124	(S) CGCTTCAC	-2,6596			(S) GTCAGTGG	-2,6768		
(E) CGAACCAAC	2,6461	(S) AGTGTCA	-1,8395	(S) CGGGCAAC	-2,5583			(S) GTCAGGGG	-3,1088		
(E) CGAACCAA	2,2048	(S) AGTGTTA	-1,6464	(S) CGTTGCGC	-2,5450			(S) GTCCACGG	-2,6020		
(E) CGAACCCG	2,6255	(S) ATACGGT	-2,0984	(S) CGTTGGCG	-2,5718			(S) GTCCGGGA	-3,4271		
(E) CGACACCC	2,3115	(S) ATATGCT	-1,7400	(S) CTCGCGCG	-2,2610			(S) GTCGGGGT	-3,4442		
(E) CGACTACG	2,1381	(S) ATCAGTG	-1,6200	(S) CTGGTCGG	-2,4990			(S) GTCGGTTA	-2,7500		
(E) CGACTTCA	2,6838	(S) ATCGAGG	-1,7505	(S) CTTCCGCG	-2,7558			(S) GTCTAGGG	-3,8704		
(E) CGAGAAGA	2,0588	(S) ATCGGGT	-2,1684	(S) CTTCTGTC	-2,8379			(S) GTCTGGGA	-3,5804		
(E) CGAGATCA	2,0638	(S) ATCGGTC	-1,6982	(S) CTTGCGAA	-2,3755			(S) GTCTTGGG	-3,2065		
(E) CGAGCCAA	2,4333	(S) ATCGTAC	-1,7643	(S) CTTGTGCA	-2,5583			(S) GTGCGCGG	-2,5494		
(E) CGATACCC	2,4510	(S) ATCGTCT	-2,0086	(S) CTTGTGTG	-2,3728			(S) GTGCTTTG	-2,9113		
(E) CGATCTCC	2,1129	(S) ATGGACA	-1,7338	(S) CTTTGCCA	-2,4413			(S) GTGGCGGG	-3,6444		
(E) CGATCTTA	2,0576	(S) ATGGGGC	-1,7720	(S) GATAGGGG	-4,0593			(S) GTGGGAGG	-2,6169		
(E) CGATTACC	2,6217	(S) ATGGGGT	-2,2288	(S) GCACTCTG	-2,6485			(S) GTGGGCGC	-4,5417		
(E) CGATTATA	3,2823	(S) ATGGGTA	-2,1950	(S) GCGGGCGG	-2,3875			(S) GTGGGGAC	-3,9505		
(E) CGCAAAAC	2,1436	(S) ATGGGTC	-2,1475	(S) GCGGGGTG	-3,7842			(S) GTGGGGGA	-3,7282		
(E) CGCCAAAA	2,1799	(S) ATGGTCT	-2,8888	(S) GCTGCGGG	-3,0720			(S) GTGGGGGG	-3,7511		
(E) CGGAGACG	2,0632	(S) ATGGTGG	-2,6389	(S) GGCATTGC	-2,2642			(S) GTGGGTCA	-4,1099		
(E) CGTAGATA	2,4747	(S) ATGGTTT	-2,4461	(S) GGGGAAGC	-2,5097			(S) GTGGGTTG	-2,9247		
(E) CTACAAAA	2,3179	(S) ATTGCCA	-1,6360	(S) GGGACATA	-2,8344			(S) GTTAGGTC	-2,5279		
(E) CTACAAGG	3,1291	(S) ATTGGGT	-1,7380	(S) GGGCAGGC	-2,3223			(S) GTTCGGAA	-2,7985		
(E) CTCCACAA	2,2893	(S) ATTGGTA	-1,8249	(S) GGGCTCAG	-2,3418			(S) GTTGGGAA	-3,6131		
(E) CTTAATAC	2,2891	(S) ATTGTC	-1,9924	(S) GGGGGTCG	-2,8003			(S) GTTTGGGG	-2,5631		
(E) GAAAAAGC	2,3153	(S) CACATGT	-1,6298	(S) GGGGTGAC	-2,6820			(S) GTTTTCAC	-2,8306		
(E) GAAAAATC	2,6419	(S) CACGGGC	-1,6631	(S) GGTAAGGG	-2,5781			(S) GTTTTGGG	-2,5386		
(E) GAAACATA	2,1142	(S) CAGATGT	-1,8907	(S) GGTCGTGC	-3,1486			(S) GTTTTGTG	-2,6141		

(E)	GAAACCAC	2,6706	(S)	CAGGGGA	-1,8465	(S)	GGTCTCGG	-3,1447			(S)	TAGGGGGG	-3,3731		
(E)	GAAACCGA	2,1855	(S)	CAGGTCG	-2,2280	(S)	GGTGGGTG	-2,7676			(S)	TCTAGGGC	-3,4164		
(E)	GAAATATC	2,9315	(S)	CAGGTTG	-1,7748	(S)	GGTTCGGC	-4,2158			(S)	TGAGGGTG	-2,6968		
(E)	GAAATCGA	2,0607	(S)	CATGAGG	-1,8605	(S)	GTA CTGCG	-2,4025			(S)	TGATCCGC	-2,7712		
(E)	GAACAGAG	2,1627	(S)	CATGCAG	-1,8440	(S)	GTATGTGC	-3,4353			(S)	TGATGTGC	-2,5958		
(E)	GAACACGA	3,0322	(S)	CATGGAC	-1,6609	(S)	GTCGGGAC	-3,8848			(S)	TGCCAGGG	-2,9795		
(E)	GAACCCCTG	2,2494	(S)	CATGGTA	-1,9629	(S)	GTCGTGTC	-2,7245			(S)	TGCCTGGC	-2,5657		
(E)	GAACCGCA	2,2860	(S)	CCACGGT	-1,6082	(S)	GTCGTGTA	-2,3806			(S)	TGCGTCCC	-2,6527		
(E)	GAACCGCC	2,6214	(S)	CCACGTG	-2,1732	(S)	GTCGGTGT	-2,3102			(S)	TGCTGCGG	-3,2732		
(E)	GAACCTCC	2,2685	(S)	CCAGGAG	-1,9218	(S)	GTCGTGCT	-2,6687			(S)	TGGAGGGG	-3,1082		
(E)	GAACCTTA	2,1977	(S)	CCAGGCG	-1,8616	(S)	GTTAGGTG	-3,1797			(S)	TGGCGGGG	-2,4474		
(E)	GAAGAAAG	3,5232	(S)	CCAGGGT	-1,6313	(S)	GTTCTGGC	-2,7494			(S)	TGGGGGTA	-2,5052		
(E)	GAAGAATA	2,1322	(S)	CCCCTGC	-1,7973	(S)	GTTGGGTA	-4,3127			(S)	TGGGGTGG	-2,9445		
(E)	GAAGACAA	2,1862	(S)	CCCGTTA	-1,8552	(S)	GTTGGGTG	-2,7442			(S)	TGGTCTAC	-2,9712		
(E)	GAAGATAA	3,9548	(S)	CCGAGGT	-2,1678	(S)	TAATTGGG	-3,2041			(S)	TGTCTGTG	-2,5420		
(E)	GAAGATAC	2,1171	(S)	CCCGCTT	-1,7129	(S)	TGCAGGGG	-2,8306			(S)	TGTGCGGG	-2,5717		
(E)	GAAGCACA	2,5052	(S)	CCTGAGC	-1,8267	(S)	TGCCCTTC	-3,3193			(S)	TGTGGGGA	-2,9451		
(E)	GAAGTCCC	2,1957	(S)	CCTGGTA	-1,7091	(S)	TGGTGTGC	-2,2491			(S)	TGTGGGGC	-3,6314		
(E)	GAAGTCGA	2,3915	(S)	CCTTGCT	-1,6278	(S)	TTCGGTGC	-2,5562			(S)	TGTTGGAA	-2,9264		
(E)	GAAGTCTA	2,1175	(S)	CGCATTG	-2,2203	(S)	TTGTCTTC	-2,9406			(S)	TTCCGGGC	-3,2442		
(E)	GAATATCA	2,8098	(S)	CGCCATG	-1,7441	(S)	AACAGGGG	-3,5289			(S)	TTCCGGGG	-2,6097		
(E)	GAATCAGA	2,0778	(S)	CGCCTTG	-1,8753	(S)	AACGGGGA	-2,7485			(S)	TTCTGTGA	-2,4876		
(E)	GAATGAAA	2,2617	(S)	CGCTGTG	-2,0938	(S)	AACTGTGC	-2,4635			(S)	TTCTGGGG	-3,4360		
(E)	GAATGACC	2,2021	(S)	CGGATGG	-2,1523	(S)	AAGGAGGC	-2,6897			(S)	TTGGGGGC	-2,7212		
(E)	GAATTCAA	2,9310	(S)	CGTAGGC	-2,0432	(S)	AATGGGCC	-2,5914			(S)	TTGGGGTC	-2,5004		
(E)	GAATTTAG	2,1373	(S)	CGTATGT	-2,0457	(S)	AATGGGGG	-3,6091			(S)	TTTGGGGC	-2,9750		
(E)	GAATTTTG	2,6176	(S)	CGTCTGC	-1,7106	(S)	ACAACCTC	-2,3127			(S)	TTTTGTGG	-3,1320		
(E)	GACAAAAA	2,5353	(S)	CGTGGGT	-1,7995	(S)	ACAGCTGC	-2,2708			(S)	TTTTTTTC	-3,1317		
(E)	GACAAAAC	2,2418	(S)	CTACGGT	-1,9846	(S)	ACCGGGGC	-3,1070			(S)	GCCACCGC	-2,5141		
(E)	GACACCCA	2,5285	(S)	CTATCGG	-1,7727	(S)	ACCGCTGC	-3,1333			(S)	GCCAGCTC	-2,8880		
(E)	GACAGACC	2,1518	(S)	CTATTGG	-1,6919	(S)	ACCTCTCG	-2,3802			(S)	GTCGTCAG	-3,0665		
(E)	GACCAACA	2,0650	(S)	CTACAGT	-2,0010	(S)	ACGCGGGG	-2,4572			(S)	AGGGGGTA	-2,6356		
(E)	GACCACCA	2,6806	(S)	CTCAGGT	-1,9193	(S)	ACGGACGG	-2,4612			(S)	GATGGGTA	-2,6880		
(E)	GACCACGA	2,7190	(S)	CTCCCGT	-1,9317	(S)	ACGGGGAG	-2,5159			(S)	TGCCCGGG	-2,9649		
(E)	GACCCTAA	2,2177	(S)	CTCGGTC	-2,1334	(S)	ACGGGGCC	-2,9096			(S)	ACGGGGTC	-2,5866		
(E)	GACCTCAG	2,3197	(S)	CTCGGTT	-2,2036	(S)	ACGGGTGC	-3,5384			(S)	CATTAGGA	-2,7947		
(E)	GACTACCA	3,3311	(S)	CTGCTAT	-1,6696	(S)	ACGTCTGT	-3,1445			(S)	CGACGTGC	-2,5682		
(E)	GAGAATGA	2,5907	(S)	CTGACAG	-1,8357	(S)	ACTAGGGG	-3,1333			(S)	GAACGGGG	-3,3278		
(E)	GAGACCCC	3,3322	(S)	CTGGAGC	-2,2734	(S)	ACTATGGC	-2,2860			(S)	GAGGGTTG	-3,8585		
(E)	GAGACCCG	2,8186	(S)	CTGGTCG	-1,7370	(S)	ACTGGGCG	-2,2456			(S)	GCTACAGG	-3,1025		
(E)	GAGATTCC	2,0698	(S)	CTGGTCT	-2,2572	(S)	ACTTGGGG	-2,3458			(S)	GGGGATCA	-2,6832		
(E)	GAGCAATA	2,3415	(S)	CTGTGTT	-2,1157	(S)	AGAACTTG	-2,4286			(S)	GGTTCATC	-2,6445		
(E)	GAGCACAC	2,4105	(S)	CTTCGGT	-1,7191	(S)	AGAACTGG	-2,2532			(S)	GTCACTAG	-2,7098		
(E)	GAGCGCAA	2,0640	(S)	CTTGCCA	-1,8508	(S)	AGAACTTG	-2,9653			(S)	TAGGGGTA	-2,5195		
(E)	GAGCTCAA	2,1499	(S)	CTTGCGG	-1,8693	(S)	AGAGCCTC	-2,3736			(S)	TGCCCGGC	-2,7338		
(E)	GAGTCATA	2,0824	(S)	GAAGGGC	-1,9657	(S)	AGATCGCG	-2,9915			(S)	TGGAGTTA	-2,7137		
(E)	GAGTCCAC	2,5313	(S)	GAAGGGT	-2,9488	(S)	AGCAGGGC	-3,3765			(S)	TTCTCAGG	-2,7676		
(E)	GATACCGG	2,1705	(S)	GAAGGTT	-2,4015	(S)	AGCAGTGG	-2,2350			(S)	GACCAAGG	-2,5879		
(E)	GATATTAA	3,3859	(S)	GAATGTT	-1,6942	(S)	AGCCTGGG	-2,4570							
(E)	GATCAGAG	2,1119	(S)	GAATGGC	-1,7452	(S)	AGCGGGAC	-2,8764							
(E)	GATCCCCC	2,2867	(S)	GAATGGT	-2,0748	(S)	AGCGGGGA	-2,6529							
(E)	GATCCCTA	2,9985	(S)	GACGAGT	-1,9082	(S)	AGCTCAGG	-2,6536							
(E)	GATCTTCA	2,2083	(S)	GACGGAT	-1,7658	(S)	AGCTCGGA	-2,2400							
(E)	GATTAGGA	2,3213	(S)	GACGGCC	-1,7602	(S)	AGCTTGTG	-2,4516							
(E)	GATTCAAA	2,7539	(S)	GACTGGT	-1,7719	(S)	AGGAGGGA	-2,5598							
(E)	GATTGACA	3,0462	(S)	GAGAGGT	-1,7679	(S)	AGGCAAGG	-2,3106							
(E)	GCAAAAAA	2,2402	(S)	GAGGCAA	-1,6890	(S)	AGGCCAGG	-2,3049							
(E)	GCAAAATA	2,4157	(S)	GAGGGCA	-2,1814	(S)	AGGCCGGA	-2,9511							
(E)	GCAACACA	2,7010	(S)	GAGGGGC	-1,6781	(S)	AGGCCGGG	-2,7076							
(E)	GCAAAACC	2,6456	(S)	GAGGGGT	-2,1340	(S)	AGGGAGTC	-3,6038							
(E)	GCAAAATTG	2,8264	(S)	GAGGGTA	-2,0803	(S)	AGGGCGGG	-2,6030							
(E)	GCAACAAC	2,6657	(S)	GAGGGTC	-1,9385	(S)	AGGGGAAC	-2,3975							
(E)	GCAACAAG	2,6159	(S)	GAGGTCG	-3,1087	(S)	AGGGGAGC	-3,2474							
(E)	GCAACCAG	2,2010	(S)	GAGGTGG	-2,2250	(S)	AGGGGGCG	-2,9581							
(E)	GCAAGAAA	2,5009	(S)	GAGTGCT	-1,8236	(S)	AGGGGTGG	-2,3089							
(E)	GCAATGTA	2,3865	(S)	GATACGT	-1,7600	(S)	AGGGGTTC	-2,6326							
(E)	GCACAACC	2,1371	(S)	GATAGGT	-2,7665	(S)	AGGGTTAG	-2,8152							
(E)	GCACACAA	2,2293	(S)	GATCGCT	-1,6653	(S)	AGGGTTGA	-2,4100							
(E)	GCACACTC	2,1049	(S)	GATGGGT	-2,7650	(S)	AGGTCGGC	-3,5109							
(E)	GCACCATA	2,4333	(S)	GATGGTA	-2,0042	(S)	AGGTGCCG	-2,8700							
(E)	GCAGAGCG	3,9014	(S)	GATGTGC	-2,8736	(S)	AGGTGTGG	-2,5374							
(E)	GCAGCTCG	2,6147	(S)	GATGTTT	-1,6392	(S)	AGTAACGG	-2,3296							
(E)	GCCAAAAC	2,9307	(S)	GCACGGT	-2,2049	(S)	AGTACGGG	-2,8031							
(E)	GCCACGAA	2,0746	(S)	GCACGCT	-1,6679	(S)	AGTACTGG	-2,7447							
(E)	GCCCGAAA	2,3864	(S)	GCAGGTT	-1,8187	(S)	AGTAGCCG	-2,6997							
(E)	GCCGAAAG	2,1464	(S)	GCAGTTT	-1,8471	(S)	AGTATGCG	-3,1628							
(E)	GCCGAGAA	2,5642	(S)	GCCAGGT	-2,1163	(S)	AGTCGGCC	-2,3784							
(E)	GCCGCGGA	2,2426	(S)	GCCAGTT	-2,3516	(S)	AGTCGGTG	-2,2612							
(E)	GCGAACAC	2,2588	(S)	GCCCCGT	-2,0204	(S)	AGTGACGG	-2,9748							
(E)	GCGAACAG	2,4499	(S)	GCCGGGT	-1,6166	(S)	AGTGCGGG	-2,3336							
(E)	GCGAACTA	2,9987	(S)	GCGAGGT	-3,0287	(S)	AGTGGCTG	-2,9267							

(E)	GCGACCAA	2,4485	(S)	GCGATTG	-1,9034	(S)	AGTGGGGG	-2,7074								
(E)	GCGACTAG	2,2802	(S)	GCGGGCA	-1,9457	(S)	AGTGTGCA	-2,9713								
(E)	GCGAGCCA	2,1346	(S)	GCGGGTA	-2,7001	(S)	AGTGGTTG	-2,4018								
(E)	GCGATCAA	2,3012	(S)	GCGGTGT	-3,2108	(S)	AGTTGCGG	-2,6214								
(E)	GCGATTCA	2,4038	(S)	GCGTGCG	-1,9880	(S)	AGTTGTAC	-2,2937								
(E)	GCGCAAAG	2,5208	(S)	GCTAAAT	-1,8990	(S)	AGTTTCAG	-2,5657								
(E)	GCGCAACC	2,3054	(S)	GCTAGGT	-2,1423	(S)	ATATGGGG	-3,0553								
(E)	GCGCACCA	2,5768	(S)	GCTCAAC	-1,8786	(S)	ATATGGTC	-2,7031								
(E)	GCGCCGAC	2,3033	(S)	GCTCAAG	-1,6508	(S)	ATATGTTG	-2,2784								
(E)	GCGGACTA	2,3911	(S)	GCTCCCT	-2,0893	(S)	ATCAAAGG	-2,2997								
(E)	GCGTACTG	2,5961	(S)	GCTCCTC	-2,1784	(S)	ATCAGCGC	-2,3969								
(E)	GCGTAGGC	2,1199	(S)	GCTGCCA	-1,6733	(S)	ATCAGGGC	-2,7004								
(E)	GCGTATTA	2,4378	(S)	GCTGGGT	-1,8507	(S)	ATCGCGGC	-2,6962								
(E)	GCGTTTAG	2,0576	(S)	GCTGTGC	-2,5254	(S)	ATCGTGGC	-2,5114								
(E)	GCTAAJAC	2,2984	(S)	GCTGTTC	-1,8269	(S)	ATCTCGGC	-2,2451								
(E)	GCTAAAGA	2,3238	(S)	GCTTCGG	-1,6404	(S)	ATCTGCCG	-2,3588								
(E)	GCTTACCA	2,2665	(S)	GCTTCCT	-1,6115	(S)	ATCTGGTC	-2,3858								
(E)	GGAAAACA	2,3358	(S)	GGAAGGT	-2,0316	(S)	ATCTTGTG	-2,4060								
(E)	GGAAAATA	2,0766	(S)	GGACACT	-1,7785	(S)	ATGCAGGG	-2,7673								
(E)	GGAACCAA	2,3641	(S)	GGACGGT	-3,0366	(S)	ATGCGGGA	-2,5293								
(E)	GGAGAGGA	2,4541	(S)	GGAGCCC	-1,6554	(S)	ATGCGGGG	-2,7327								
(E)	GGACCCTA	2,1915	(S)	GGAGGTA	-2,5883	(S)	ATGGCATC	-2,3886								
(E)	GGAGCCAA	2,3755	(S)	GGAGTAC	-2,4098	(S)	ATGGGAGC	-2,5268								
(E)	GGATACAA	2,8701	(S)	GGAGTCC	-1,6156	(S)	ATGGGGAA	-3,7911								
(E)	GGATCACC	2,1530	(S)	GGAGTGC	-2,0409	(S)	ATGGGGTC	-2,3846								
(E)	GGCCCAACA	2,1002	(S)	GGATCCC	-2,0267	(S)	ATGGGTAA	-2,2350								
(E)	GTACACCC	2,3181	(S)	GGCAAGT	-2,3880	(S)	ATGGGTGA	-3,2821								
(E)	GTCGAAAC	2,1184	(S)	GGCACAC	-1,9304	(S)	ATGGTGCA	-2,3389								
(E)	GTCGACCC	2,4256	(S)	GGCACGA	-1,6820	(S)	ATGGTGGG	-2,4557								
(E)	GTTGAAGA	2,2700	(S)	GGCACGT	-2,3521	(S)	ATGTCTGG	-2,6459								
(E)	TAAAAACA	3,9157	(S)	GGCAGCC	-1,6350	(S)	ATGTGATG	-3,0491								
(E)	TAAACATA	2,6056	(S)	GGCAGGT	-2,5011	(S)	ATGTGGTG	-2,2783								
(E)	TAAACCAA	2,9056	(S)	GGCGAAC	-1,8477	(S)	ATGTGTGG	-2,8450								
(E)	TAAACCGA	2,4595	(S)	GGCAGGT	-1,6384	(S)	ATGTTTGC	-2,2969								
(E)	TAAACTCC	2,1223	(S)	GGCGGCA	-1,7329	(S)	ATTCCGTC	-2,7455								
(E)	TAAACTGA	2,6380	(S)	GGCGGGT	-1,8606	(S)	ATTGGGGC	-2,5487								
(E)	TAAAGACA	2,1820	(S)	GGGACAC	-1,7294	(S)	ATTGGGGG	-2,4917								
(E)	TAAATCAA	2,0983	(S)	GGGACCC	-1,6495	(S)	ATTGGTGC	-2,4503								
(E)	TAAATTCA	2,1192	(S)	GGGACTC	-1,7914	(S)	ATTGGTTG	-2,6150								
(E)	TACAAAC	2,6065	(S)	GGGAGGT	-2,9616	(S)	CAAAGGGA	-3,3693								
(E)	TACAACAG	2,3784	(S)	GGGATAT	-1,7636	(S)	CAAGCGGG	-2,8600								
(E)	TACACAAA	2,1243	(S)	GGGCGAT	-1,8194	(S)	CAAGGCTC	-2,7907								
(E)	TACACACA	3,1058	(S)	GGGCGTA	-1,6293	(S)	CAAGGTGC	-3,9819								
(E)	TACAGAGC	2,6590	(S)	GGGGGTA	-2,7166	(S)	CAATGGGG	-2,9019								
(E)	TACATAAA	2,6134	(S)	GGGGTAT	-3,1196	(S)	CAATTACG	-2,7041								
(E)	TACCAAC	2,3632	(S)	GGGGTCT	-2,5328	(S)	CACAGGCA	-3,0750								
(E)	TACCCCTA	2,2817	(S)	GGGTAAC	-1,7887	(S)	CACAGGGG	-2,9347								
(E)	TACCCCTC	2,0613	(S)	GGGTAAT	-1,7347	(S)	CACAGTGG	-2,3419								
(E)	TACCTTAC	2,2860	(S)	GGGTACA	-2,0149	(S)	CACATGGG	-2,4738								
(E)	TACCGAAG	2,3642	(S)	GGGTACC	-3,2363	(S)	CACCGGGG	-2,6328								
(E)	TACCTTAA	2,7420	(S)	GGGTACT	-2,1131	(S)	CACCGCTC	-2,3802								
(E)	TAGAGAAAG	3,8380	(S)	GGGTATC	-2,6783	(S)	CACCGGGG	-2,5351								
(E)	TAGAGCCA	2,4603	(S)	GGGTATT	-1,7305	(S)	CACGAGGG	-2,8382								
(E)	TAGAGCCC	2,9496	(S)	GGGTCTC	-2,2491	(S)	CACGCCTC	-2,2941								
(E)	TAGAGCTA	2,4044	(S)	GGTAACG	-1,9910	(S)	CACGCGTG	-3,6097								
(E)	TAGAGAGG	2,5607	(S)	GGTAAGT	-2,0610	(S)	CACGCGCG	-2,2345								
(E)	TAGATATA	2,6859	(S)	GGTAAGG	-3,4802	(S)	CACGGGGA	-2,6327								
(E)	TAGACACG	2,5415	(S)	GGTACAA	-1,8250	(S)	CACGGTGC	-2,2461								
(E)	TAGACCCA	3,1048	(S)	GGTACAT	-1,7806	(S)	CACGTAGC	-3,3334								
(E)	TAGAGCTCA	2,4904	(S)	GGTACCG	-2,2346	(S)	CACTCGGG	-3,2693								
(E)	TAGAGTGA	2,8189	(S)	GGTACGG	-2,6354	(S)	CACTCGGG	-2,7413								
(E)	TAGGAAC	2,1449	(S)	GGTACTG	-2,0422	(S)	CACTCGGG	-3,0073								
(E)	TAGGTCAA	2,1041	(S)	GGTAGAA	-1,8240	(S)	CAGCAGTG	-2,3853								
(E)	TAATATCC	2,1371	(S)	GGTAGAT	-2,8221	(S)	CAGCATTC	-2,5942								
(E)	TAATATTC	2,1252	(S)	GGTAGCA	-3,1602	(S)	CAGCCGGG	-2,3371								
(E)	TAATCTAA	2,2973	(S)	GGTAGGG	-3,0318	(S)	CAGCGCGC	-2,6890								
(E)	TACAAAGA	2,6124	(S)	GGTAGGT	-4,0181	(S)	CAGCGCGG	-2,2891								
(E)	TACAACTA	2,3620	(S)	GGTATAT	-2,4435	(S)	CAGCGTGG	-2,3356								
(E)	TACAATAA	2,1255	(S)	GGTATCC	-2,9832	(S)	CAGCGTGA	-2,3773								
(E)	TACACACA	2,0674	(S)	GGTCATC	-2,5560	(S)	CAGCGTGC	-2,3965								
(E)	TACACCCA	3,3004	(S)	GGTCGGT	-3,1034	(S)	CAGCTGTA	-2,4359								
(E)	TACATTAA	2,4245	(S)	GGTCCTC	-2,9418	(S)	CAGCTTTC	-2,4482								
(E)	TACCAAAA	2,0595	(S)	GGTCGCC	-1,6815	(S)	CAGGACGG	-2,6044								
(E)	TACCAACA	2,6217	(S)	GGTCGGA	-2,5831	(S)	CAGGAGGG	-2,8716								
(E)	TACCAATA	2,4639	(S)	GGTCGGC	-3,1751	(S)	CAGGCATG	-2,8980								
(E)	TACCACGC	2,2247	(S)	GGTCGTT	-2,1158	(S)	CAGGCTAA	-2,2531								
(E)	TACCACTC	2,3761	(S)	GGTCTCC	-1,8104	(S)	CAGGGCAG	-3,3441								
(E)	TACCACACA	2,1255	(S)	GGTCTCT	-1,9189	(S)	CAGGGCGC	-3,4433								
(E)	TACCCACG	2,2320	(S)	GGTCTGC	-3,1929	(S)	CAGGGGAA	-3,4374								
(E)	TACGATCG	2,7517	(S)	GGTGAAC	-2,4232	(S)	CAGGGGAC	-2,4225								
(E)	TACGGATG	2,2885	(S)	GGTGACA	-2,1092	(S)	CAGGGGTC	-2,3932								

(E)	TACTAGAA	2,1228	(S)	GGTGACC	-2,4491	(S)	CAGGTGGC	-2,7995								
(E)	TACTCACA	2,8568	(S)	GGTGACT	-2,3923	(S)	CAGTACGC	-2,3015								
(E)	TACTCAGA	2,0634	(S)	GGTGAGG	-2,1064	(S)	CAGTCCTC	-2,2989								
(E)	TACTCCGA	2,4295	(S)	GGTGATA	-1,9799	(S)	CAGTCGGG	-2,3666								
(E)	TACTCTAA	2,3595	(S)	GGTGATT	-2,1649	(S)	CAGTGCGA	-2,2479								
(E)	TACTTACC	2,8059	(S)	GGTGCCC	-2,1788	(S)	CAGTGCGC	-2,2538								
(E)	TAGAAATA	2,8831	(S)	GGTGCTC	-2,4681	(S)	CAGTGTGC	-2,4394								
(E)	TAGAACAG	2,5940	(S)	GGTGCTT	-3,2169	(S)	CAGTTTGG	-2,6618								
(E)	TAGAAGAC	2,3743	(S)	GGTGCCA	-1,9705	(S)	CATCGGTC	-2,5075								
(E)	TAGAAGCA	2,1129	(S)	GGTGGCC	-2,7939	(S)	CATGAGGC	-2,4205								
(E)	TAGACAAG	2,3922	(S)	GGTGGTC	-3,8206	(S)	CATGCCGG	-2,3555								
(E)	TAGACATC	3,0909	(S)	GGTGTTT	-2,7792	(S)	CATGGCAG	-3,0402								
(E)	TAGACCCA	3,2588	(S)	GGTGTAC	-3,6808	(S)	CATGGGGA	-4,1594								
(E)	TAGACCTA	2,9211	(S)	GGGTAT	-1,6739	(S)	CATGGGTC	-2,5715								
(E)	TAGACCTG	3,3418	(S)	GGGTGCA	-2,6830	(S)	CATGGTTC	-2,6651								
(E)	TAGACTGG	2,1030	(S)	GGGTGCC	-2,1076	(S)	CATGTGTA	-2,4332								
(E)	TAGATCGA	2,7339	(S)	GGGTGT	-4,0571	(S)	CATTCTGG	-2,2767								
(E)	TAGATTAC	2,2076	(S)	GGGTGTC	-2,9504	(S)	CATTGGGA	-2,3704								
(E)	TAGCAACA	2,2591	(S)	GGTTCGC	-2,0951	(S)	CATTGTCA	-3,1374								
(E)	TAGCAATA	2,3817	(S)	GGTTGCC	-2,7159	(S)	CCACGGGG	-4,6675								
(E)	TAGCACAG	2,5309	(S)	GTACCCG	-2,1611	(S)	CCAGCCCG	-2,6788								
(E)	TAGCACCA	2,3845	(S)	GTACGGC	-1,8359	(S)	CCAGGGAG	-3,7327								
(E)	TAGCCAAA	2,5724	(S)	GTACGTT	-2,0207	(S)	CCAGTCTG	-2,3104								
(E)	TAGCCAGA	2,2251	(S)	GTATCGA	-1,8725	(S)	CCCAGGGC	-2,6530								
(E)	TAGCGCTA	2,5313	(S)	GTATGGT	-2,5629	(S)	CCCCCGTG	-2,3197								
(E)	TAGCGGCA	2,1637	(S)	GTATGTT	-2,5499	(S)	CCCCGGCG	-2,5831								
(E)	TAGCTGTG	2,0640	(S)	GTCAGGT	-2,4619	(S)	CCCGCAGC	-2,2903								
(E)	TAGGAACA	2,3317	(S)	GTCAGTC	-1,6923	(S)	CCCGCCGG	-2,5497								
(E)	TAGGACAA	2,1705	(S)	GTCGGAT	-2,1084	(S)	CCCGCGTG	-2,6980								
(E)	TAGGACTA	2,6173	(S)	GTCGAGT	-1,7887	(S)	CCCGGCGC	-2,6427								
(E)	TAGGATCA	2,3775	(S)	GTCGGGT	-2,4845	(S)	CCCGGCGG	-2,2799								
(E)	TAGGCACC	2,1708	(S)	GTCCTGC	-1,6926	(S)	CCCGGGCG	-2,6404								
(E)	TAGGTACA	2,2015	(S)	GTGCCGG	-1,7850	(S)	CCCGGGGA	-3,1236								
(E)	TAGGTCAA	2,3706	(S)	GTGGTTC	-2,1483	(S)	CCCGGTCC	-2,7301								
(E)	TAGTAGCA	2,2981	(S)	GTGCGTT	-2,0651	(S)	CCCGGTGC	-4,1020								
(E)	TAGTCAGA	2,5435	(S)	GTGGAAG	-1,6112	(S)	CCCGTGGC	-2,4433								
(E)	TAGTCCAC	2,5396	(S)	GTGGGTA	-3,0395	(S)	CCCGTTGG	-2,7385								
(E)	TAGTGACA	3,0363	(S)	GTGGTCC	-3,4797	(S)	CCCTATGA	-2,4836								
(E)	TATAAACA	2,2564	(S)	GTGGTGT	-2,5231	(S)	CCCTCGCA	-2,9411								
(E)	TATAAGAG	2,3141	(S)	GTGTCGA	-2,0065	(S)	CCCTGGGG	-2,9041								
(E)	TATAAGTA	2,1900	(S)	GTGTGAC	-1,6924	(S)	CCCTGTCA	-2,3767								
(E)	TATACTAC	2,0732	(S)	GTTAGGT	-2,6470	(S)	CCCTTCGC	-2,2994								
(E)	TATCCACA	2,1483	(S)	GTTATGC	-1,9974	(S)	CCGAGGGG	-2,4709								
(E)	TATCCTCA	2,9529	(S)	GTTCCCC	-2,3687	(S)	CCGCGGGC	-3,2572								
(E)	TATGACCC	2,4857	(S)	GTTGGGT	-2,4047	(S)	CCGCTTTC	-2,7221								
(E)	TATGATGA	2,2782	(S)	GTTGTAC	-2,4813	(S)	CCGGCCAA	-2,6511								
(E)	TATGCACC	2,1748	(S)	GTTGTGC	-1,8936	(S)	CCGGGGAA	-3,2132								
(E)	TATTAACA	2,6159	(S)	GTTTTC	-1,7154	(S)	CCGGGGGC	-2,3699								
(E)	TATTAACG	2,3455	(S)	TAACGTT	-1,7566	(S)	CCGGTGGG	-3,7407								
(E)	TATTAATA	2,5048	(S)	TAGCATT	-1,9323	(S)	CCGTATGC	-2,3617								
(E)	TATTATCA	2,2408	(S)	TAGGGAC	-1,7057	(S)	CCGTGCGC	-2,8251								
(E)	TATTCCAA	2,4345	(S)	TAGGGCA	-1,6254	(S)	CCGTGCTG	-2,2843								
(E)	TCAAAACG	2,6111	(S)	TAGGGGA	-1,7521	(S)	CCGTGTAG	-2,7721								
(E)	TCAAAACG	2,5566	(S)	TAGGGGT	-2,6644	(S)	CCGTGTGC	-2,4621								
(E)	TCAACAAG	2,2642	(S)	TAGGGTC	-1,9669	(S)	CCGTGTGG	-2,3633								
(E)	TCAACCAC	2,3269	(S)	TATCGGG	-1,6409	(S)	CCGTTACG	-2,8066								
(E)	TCAACGCC	2,1401	(S)	TATGTAC	-1,6406	(S)	CCGTTATC	-2,4970								
(E)	TCAAGAAG	2,1470	(S)	TCAGGTT	-1,8697	(S)	CCTAGGCC	-2,3334								
(E)	TCAAGACC	3,1292	(S)	TCGGACG	-1,6948	(S)	CCTATACG	-2,3394								
(E)	TCAAGCCA	3,3923	(S)	TCGGGAC	-1,7582	(S)	CCTCCGGG	-3,8222								
(E)	TCACAAGG	2,0845	(S)	TCGGGGT	-1,8539	(S)	CCTCGGTG	-3,3041								
(E)	TCACACCA	3,1803	(S)	TCTCGTC	-1,6445	(S)	CCTCGGTA	-4,2638								
(E)	TCACACTA	2,5586	(S)	TCTGTGC	-1,8389	(S)	CCTCTTTG	-2,3716								
(E)	TCACCCCG	2,8965	(S)	TCTTGGC	-1,7555	(S)	CCTGCACG	-2,6519								
(E)	TCACTCCA	3,5153	(S)	TGAGGTT	-2,6124	(S)	CCTGCCGG	-2,4662								
(E)	TCAGATAA	2,1334	(S)	TGCATT	-1,9307	(S)	CCTGCCTC	-3,3319								
(E)	TCAGACCA	2,4202	(S)	TGCGCTT	-1,8703	(S)	CCTGCGGA	-2,4886								
(E)	TCAGCCCG	2,4988	(S)	TGCTTTA	-2,0372	(S)	CCTGGATC	-2,4381								
(E)	TCAGCTAA	2,2748	(S)	TGGAGGT	-2,5281	(S)	CCTGGGGC	-2,5883								
(E)	TCAGCTCG	2,4785	(S)	TGGGGCT	-1,7868	(S)	CCTGTGCG	-2,4430								
(E)	TCAGGAAC	2,5157	(S)	TGGGGTA	-2,5578	(S)	CCTTCAGG	-2,3429								
(E)	TCAGTCAG	2,1393	(S)	TGGGGTC	-2,1775	(S)	CCTTGAC	-2,5100								
(E)	TCATAATG	2,1403	(S)	TGGGTCT	-2,0822	(S)	CCTTGGGG	-3,1168								
(E)	TCATATCC	2,1963	(S)	TGGGTGG	-1,8300	(S)	CCTTGTGG	-2,4441								
(E)	TCATCCCC	2,1493	(S)	TGTAGGC	-1,9681	(S)	CGAAAGGG	-3,5405								
(E)	TCATCCCG	2,1448	(S)	TGTAGGT	-1,9538	(S)	CGACATCG	-2,4035								
(E)	TCATTCCA	2,9070	(S)	TGTATGC	-2,2849	(S)	CGACGTGG	-2,6902								
(E)	TCCAAACG	2,8800	(S)	TGTGGGC	-1,7751	(S)	CGAGGTAG	-2,9826								
(E)	TCCAACTC	2,1664	(S)	TGTGTGC	-1,6280	(S)	CGATGGGC	-2,3105								
(E)	TCCACGCA	2,2218	(S)	TTATGGT	-1,8770	(S)	CGATTGCG	-2,2989								
(E)	TCCCCAAG	2,7409	(S)	TTATTGC	-1,7363	(S)	CGCAACGG	-2,4022								

(E) TCCCCCAA	2,8534	(S) TTCATTG	-2,0043	(S) CGCACGTG	-2,8961						
(E) TCCCGAAC	2,6558	(S) TTTAGGT	-1,8136	(S) CGCAGCGA	-2,2506						
(E) TCCGAACA	3,3793	(S) TTTATGG	-1,8897	(S) CGCAGGGA	-3,1355						
(E) TCCGAAGC	2,7118			(S) CGCCATGA	-2,9071						
(E) TCCGAAGG	2,6781			(S) CGCCCGCA	-2,3246						
(E) TCCGCACA	3,2402			(S) CGCCCGGC	-2,4068						
(E) TCCGGAAG	2,1892			(S) CGCCCGTC	-2,8280						
(E) TCCGGCCA	2,3964			(S) CGCCCGAG	-2,8976						
(E) TCCGTCCA	2,4353			(S) CGCCGGGG	-3,7117						
(E) TCCTACCG	2,1265			(S) CGCCGGTG	-2,7709						
(E) TCCTTACC	2,2977			(S) CGCCGTGC	-2,4057						
(E) TCGAACTG	2,5335			(S) CGCCGTTC	-2,6656						
(E) TCGAAGAA	2,5080			(S) CGCCTGGA	-2,2466						
(E) TCGAAGTA	2,5519			(S) CGCGCGTC	-2,9645						
(E) TCGAAGTG	2,7354			(S) CGCGGGAA	-2,2383						
(E) TCGAATAG	2,4529			(S) CGCGGGGC	-2,6503						
(E) TCGAATTA	3,7287			(S) CGCGGGTC	-2,3537						
(E) TCGACAAA	2,5630			(S) CGCGGTAG	-2,3683						
(E) TCGACAAAC	2,6206			(S) CGCGGTCC	-2,2499						
(E) TCGACCGC	2,0939			(S) CGCGGTGG	-2,4697						
(E) TCGACGCC	2,2470			(S) CGCGGTTC	-2,7010						
(E) TCGAGAGG	2,1366			(S) CGCGTGGG	-2,6363						
(E) TCGAGATA	2,3092			(S) CGCGGTGC	-3,1776						
(E) TCGAGCCG	3,1263			(S) CGCGTTCC	-2,5080						
(E) TCGATACG	2,5004			(S) CGCGTTGC	-3,4362						
(E) TCGCACAG	2,7329			(S) CGCTAGCC	-2,4986						
(E) TCGCAGTC	2,6796			(S) CGCTCCGC	-2,7000						
(E) TCGCAGTG	2,8228			(S) CGCTCTTG	-3,7480						
(E) TCGCCCTC	2,0558			(S) CGCTGTGA	-2,3418						
(E) TCGCCTAC	2,7007			(S) CGCTTCTG	-3,6962						
(E) TCGCCTAG	2,6560			(S) CGCTTTGC	-3,0439						
(E) TCGCGTCC	2,1265			(S) CGCTTTGC	-2,8888						
(E) TCGCGTGG	2,0958			(S) CGGAACGC	-2,4464						
(E) TCGCTATA	2,4733			(S) CGGAGCAC	-2,4311						
(E) TCGCTCAA	2,2141			(S) CGGAGGGA	-3,4807						
(E) TCGCTCAC	2,2856			(S) CGGAGGGC	-2,5507						
(E) TCGCTCGA	2,5436			(S) CGGCCGTC	-3,4624						
(E) TCGCTGTA	2,2810			(S) CGGCCTTG	-2,8408						
(E) TCGGACAA	2,7843			(S) CGGCGGGC	-2,9520						
(E) TCGGACGG	2,4609			(S) CGGCGGGG	-2,6386						
(E) TCGGAGGA	2,0802			(S) CGGAGGCC	-2,4588						
(E) TCGGATAG	3,0542			(S) CGGGCGGG	-2,3025						
(E) TCGGATCA	2,6657			(S) CGGGCTTA	-2,2835						
(E) TCGGATCC	2,2775			(S) CGGGGAGA	-3,2608						
(E) TCGGCAGC	2,5162			(S) CGGGGAGC	-4,1500						
(E) TCGGCAGG	2,5541			(S) CGGGGAGG	-4,5044						
(E) TCGGCATC	2,2432			(S) CGGGGATG	-2,4424						
(E) TCGGCCAC	2,4189			(S) CGGGGCTC	-2,4703						
(E) TCGGCCTA	2,6824			(S) CGGGGGAC	-2,6257						
(E) TCGGCGAG	2,2069			(S) CGGGGGTA	-3,4451						
(E) TCGGCGGC	2,3444			(S) CGGGGTCC	-2,2800						
(E) TCGGGCGC	2,1789			(S) CGGGGTGC	-3,8545						
(E) TCGGGGCA	2,6173			(S) CGGGGTTC	-2,2977						
(E) TCGGTACA	2,5084			(S) CGGGTGGG	-3,1187						
(E) TCGGTGCG	2,0584			(S) CGGGTTGG	-2,4686						
(E) TCGTACTG	3,0878			(S) CGGTACAC	-2,2813						
(E) TCGTAGAC	2,5249			(S) CGGTCTGC	-2,3632						
(E) TCGTAGGA	2,5336			(S) CGGTCTGG	-2,3738						
(E) TCGTCCTA	2,5343			(S) CGGTGATC	-2,3863						
(E) TCGTCGAC	3,1834			(S) CGGTGCAG	-2,4501						
(E) TCGTCGGG	2,4424			(S) CGGTGCGG	-2,6224						
(E) TCGTGAAG	2,1593			(S) CGGTGTGC	-2,8865						
(E) TCGTGAGA	2,4212			(S) CGGTGCC	-2,9493						
(E) TCGTGATA	2,9378			(S) CGGTTTTT	-3,4794						
(E) TCGTGACAC	3,0171			(S) CGTAGTAG	-2,8687						
(E) TCGTTGTG	2,4210			(S) CGTATCTC	-3,2235						
(E) TCTAACCA	2,6859			(S) CGTCATGC	-2,7766						
(E) TCTAACAC	2,4516			(S) CGTCGCGC	-2,6211						
(E) TCTACATC	2,1803			(S) CGTCGGCC	-4,3654						
(E) TCTAGAAG	2,7088			(S) CGTCGGGA	-2,7683						
(E) TCTCAATC	2,2656			(S) CGTCGGGC	-2,4094						
(E) TCTGACGA	2,4629			(S) CGTCGGGG	-3,0953						
(E) TCTGACTG	2,0740			(S) CGTCGGTC	-2,5600						
(E) TCTGGAAA	2,1634			(S) CGTCTCGG	-2,7086						
(E) TGAACACC	2,0714			(S) CGTGGCTG	-2,7848						
(E) TGAACCAA	2,4633			(S) CGTGGGGC	-4,3010						
(E) TGAATCAA	2,8530			(S) CGTGGGTC	-2,7124						
(E) TGAATCCC	2,0625			(S) CGTGGTCG	-3,3773						
(E) TGACAACA	2,7270			(S) CGTGGTGA	-2,3549						
(E) TGACCATC	2,2075			(S) CGTGGTGC	-2,2386						
(E) TGACTTCC	2,3846			(S) CGTGTCTC	-2,3036						

(E) TGAGAAAG	2,1514		(S) CGTGTGCC	-2,2435						
(E) TGAGAATA	2,3077		(S) CGTTCAGA	-2,4721						
(E) TGAGACCC	2,5755		(S) CGTTCAGC	-3,2472						
(E) TGAGAGAC	2,4233		(S) CGTTGCGC	-2,6832						
(E) TGATAAAA	3,0808		(S) CGTTGCGG	-3,2688						
(E) TGATCCCG	2,2766		(S) CGTTGGAC	-3,2698						
(E) TGCGAAAA	2,3404		(S) CGTTGGGC	-2,5907						
(E) TGCGAGCA	2,3768		(S) CGTTGGGG	-2,8726						
(E) TGGGACAG	2,2445		(S) CGTTTTGC	-2,5301						
(E) TGGGACCA	2,6315		(S) CGTTTTGG	-2,3680						
(E) TGGTAGAA	2,3163		(S) CTACGGGC	-3,3559						
(E) TGTAACCG	2,1102		(S) CTACGGTC	-2,8894						
(E) TTAATCCC	2,1546		(S) CTACGTAG	-3,1444						
(E) TTATAATA	2,7083		(S) CTACTCGG	-2,4939						
(E) TTCAAACA	2,5168		(S) CTACTGGG	-2,4276						
(E) TTCGAGCA	2,5356		(S) CTATCAGC	-2,5391						
(E) TTCGCCGA	2,2186		(S) CTATCAGG	-2,4598						
(E) TTGCAACG	2,2052		(S) CTATGTAG	-2,3197						
(E) TTGCAATA	2,7762		(S) CTATGTGA	-2,5545						
(E) TTGCACCC	3,5373		(S) CTATTCGG	-2,5641						
(E) TTGCCAGA	2,3032		(S) CTATTTTA	-2,3606						
(E) TTGCCCGA	2,8841		(S) CTCAGCGC	-2,4327						
(E) TTGCCGAC	2,6121		(S) CTCAGGTC	-2,7024						
(E) TTGGGGCA	2,0826		(S) CTCAGTGA	-2,3908						
(E) TTGTAGAA	2,3904		(S) CTCAGATC	-2,3491						
(E) TTGTGACG	2,4303		(S) CTCGGTTG	-2,2936						
(E) TTTACCCG	2,2999		(S) CTCGAGGG	-2,2474						
(E) TTTACGCC	2,1299		(S) CTCGCCGG	-2,5535						
(E) TTTACTAA	2,7842		(S) CTCGCGCG	-2,5426						
(E) TTTTAAAA	2,4679		(S) CTCGCCGC	-2,9789						
(S) AAAGGGCG	-2,3353		(S) CTCGCCGG	-2,7350						
(S) AACGCGGC	-2,7778		(S) CTCGCGTG	-2,9916						
(S) AACGTGTC	-2,3496		(S) CTCGCTGG	-2,7245						
(S) AAGCTGTC	-2,2105		(S) CTCGGGCC	-2,3274						
(S) AAGGGGAA	-2,2323		(S) CTCGGTAA	-2,5882						
(S) AAGGGGTA	-2,8633		(S) CTCGGTCG	-2,5964						
(S) AAGGGTGG	-2,2587		(S) CTCGGTGA	-2,2755						
(S) AAGGGTTC	-3,3894		(S) CTCGTTCG	-2,6186						
(S) AAGTGGGG	-2,9684		(S) CTCTGATG	-3,1078						
(S) AATCGGGG	-2,5041		(S) CTCTGGGC	-3,2170						
(S) AATGGGGA	-2,1243		(S) CTCTGGGG	-2,6579						
(S) AATTTGGG	-2,1441		(S) CTCTGTGG	-3,9098						
(S) ACACGGAA	-2,0642		(S) CTCTTGGC	-2,5298						
(S) ACACGGGG	-2,7005		(S) CTGCGCAC	-2,7056						
(S) ACAGGGGC	-3,2842		(S) CTGCGCCA	-2,3587						
(S) ACAGGTGC	-2,2240		(S) CTGTGGGC	-3,0130						
(S) ACAGTGGG	-2,2671		(S) CTGCTGGG	-2,6671						
(S) ACATGGAC	-2,0560		(S) CTGGATTA	-2,8743						
(S) ACATGGCG	-2,1055		(S) CTGGCAAC	-2,3069						
(S) ACCTGGGG	-2,7596		(S) CTGGCAGG	-2,3721						
(S) ACGACGGC	-2,1222		(S) CTGGCCGC	-2,4431						
(S) ACGGCTCG	-2,1689		(S) CTGGCTGC	-2,5817						
(S) ACGGGGCG	-2,4443		(S) CTGGGAAA	-3,6153						
(S) ACGGGGGG	-2,0686		(S) CTGGGAGC	-3,2931						
(S) ACGGGGTA	-2,1186		(S) CTGGGCGC	-3,0032						
(S) ACGGTGGG	-2,5916		(S) CTGGGTGA	-2,4392						
(S) ACGTGAGG	-2,6597		(S) CTGGGTTC	-2,5666						
(S) ACGTGGGC	-2,7347		(S) CTGGTCGA	-2,8843						
(S) ACTCGGTC	-2,3375		(S) CTGGTGGC	-3,1249						
(S) ACTCGGTC	-2,5230		(S) CTGTACGC	-2,2699						
(S) ACTGATGG	-2,1869		(S) CTGTCCGG	-2,5996						
(S) ACTGCGTC	-2,7082		(S) CTGTCCTG	-2,2792						
(S) ACTGGGAC	-2,2557		(S) CTGTGAGG	-2,7074						
(S) ACTGTGCG	-2,6191		(S) CTGTGCAC	-2,4687						
(S) ACTTGGGA	-2,1518		(S) CTGTGGTA	-3,8511						
(S) AGACGCGC	-2,2878		(S) CTGTGTTC	-3,1033						
(S) AGACGGGC	-2,1624		(S) CTGTTATG	-2,8893						
(S) AGATCTTC	-2,4012		(S) CTGTTGGG	-3,2185						
(S) AGATGGTG	-2,0934		(S) CTGTTGGG	-2,8009						
(S) AGCACTTC	-2,4889		(S) CTTAGTAG	-3,3079						
(S) AGCAGAGG	-2,3151		(S) CTTAGTGG	-2,4819						
(S) AGCATCGC	-2,1438		(S) CTTATGGC	-2,4601						
(S) AGCCCTCG	-2,1088		(S) CTTCCGGG	-3,0579						
(S) AGCCCGTG	-2,0724		(S) CTTCCGCA	-3,4474						
(S) AGCGAGGC	-2,4305		(S) CTTCCGGC	-2,7298						
(S) AGCGCTCG	-2,5198		(S) CTTCGTTA	-3,8083						
(S) AGCGGCCC	-2,8618		(S) CTTCTTGC	-2,4337						
(S) AGCGGCGG	-2,5725		(S) CTTGAGCG	-2,3150						
(S) AGCGGGCC	-3,0504		(S) CTTGCGCG	-3,0956						
(S) AGCGGGTC	-2,4285		(S) CTTGCGGG	-2,3184						
(S) AGCGGGTG	-2,1027		(S) CTTGCGTG	-2,5635						

(S) AGCGTGGG	-2,1352		(S) CTTGCTGG	-2,6908						
(S) AGCTAGGC	-2,9084		(S) CTTGGATA	-2,9499						
(S) AGGAAAGC	-2,4234		(S) CTTGGGAA	-2,4752						
(S) AGGACTTC	-2,1710		(S) CTTGGGCC	-2,8666						
(S) AGGAGGGG	-2,7073		(S) CTTGGGGC	-3,0480						
(S) AGGATGGC	-2,0609		(S) CTTGGGGG	-2,5795						
(S) AGGATTGG	-2,1211		(S) CTTGTGAG	-2,4737						
(S) AGGCACGC	-2,4759		(S) CTTGTGCG	-2,3373						
(S) AGGCAGTG	-2,2016		(S) CTTGTTCG	-2,4014						
(S) AGGCATGA	-2,3517		(S) CTTTCACG	-2,4974						
(S) AGGCCCGA	-2,3055		(S) CTTTCGCG	-2,2481						
(S) AGGCCCGC	-2,8140		(S) CTTTCCTG	-2,4486						
(S) AGGCCCGG	-3,2215		(S) CTTTCGCC	-3,6316						
(S) AGGCCTCG	-2,1151		(S) CTTTCGGG	-2,4502						
(S) AGGCGCTG	-2,4310		(S) CTTTCCTG	-2,7938						
(S) AGGCATGA	-3,1170		(S) CTTTCCTG	-2,3612						
(S) AGGCGGTC	-2,0909		(S) CTTTCCTG	-2,9499						
(S) AGGGGATG	-3,0269		(S) CTTTGACG	-3,1541						
(S) AGGGGGAA	-2,9229		(S) CTTTGGA	-2,5704						
(S) AGGGGGTA	-2,8818		(S) CTTTGTTG	-2,9281						
(S) AGGGTCCA	-2,1913		(S) CTTTGTGC	-2,3172						
(S) AGGTATGA	-2,6054		(S) GAAAGGGG	-2,2594						
(S) AGGTATTC	-2,1498		(S) GAAGGGAA	-2,8225						
(S) AGGTCTGG	-2,2643		(S) GAAGGGTC	-2,6380						
(S) AGGTGTTG	-2,1779		(S) GACAGTTG	-2,4948						
(S) AGGTTCGG	-2,1902		(S) GACCGCGG	-2,6404						
(S) AGTACTAG	-2,1722		(S) GACCGCGG	-2,3094						
(S) AGTAGTGC	-2,5873		(S) GACGGTTG	-2,2352						
(S) AGTCCCGG	-3,3879		(S) GACGTGGG	-2,3057						
(S) AGTCGGGG	-2,2704		(S) GAGCGCGG	-2,5976						
(S) AGTGACAG	-2,0764		(S) GAGCGGAC	-2,5914						
(S) AGTGGCGC	-2,5530		(S) GAGGGGGC	-2,4645						
(S) AGTGGGAC	-2,3942		(S) GAGGGGTG	-3,5691						
(S) AGTGGGCA	-2,2546		(S) GAGTAAGG	-2,2968						
(S) AGTGGGCC	-2,4630		(S) GATCGGGA	-2,9759						
(S) AGTGGTGG	-2,4828		(S) GATGGTCA	-2,6232						
(S) AGTGTGAG	-2,5198		(S) GATGTGGC	-3,0863						
(S) AGTTGGGC	-2,3817		(S) GATTGCGG	-2,2366						
(S) AGTTGGTA	-2,5549		(S) GCAACGGG	-2,4639						
(S) AGTTTGTG	-2,0821		(S) GCAATTGG	-3,1373						
(S) ATACTTGC	-2,1396		(S) GCACGGTC	-2,3973						
(S) ATCAGGTC	-2,2727		(S) GCACGTGC	-2,4543						
(S) ATCGCACG	-2,1988		(S) GCAGGGTC	-2,6969						
(S) ATCGGAAC	-2,2834		(S) GCAGGTGG	-2,7746						
(S) ATCGGGTA	-2,2004		(S) GCATGGGC	-2,4882						
(S) ATCGTCTA	-2,4069		(S) GCCACGTG	-2,3886						
(S) ATCGTGGG	-2,0768		(S) GCCAGGGC	-2,2343						
(S) ATCTACGG	-2,8923		(S) GCCAGTTG	-2,9781						
(S) ATCTGCTG	-2,0565		(S) GCCCGCGG	-2,4154						
(S) ATCTGGGC	-2,3169		(S) GCCCGCGG	-3,1174						
(S) ATGCGAGC	-2,2314		(S) GCCCGGAA	-2,8783						
(S) ATGCGCGG	-2,7228		(S) GCCCGTGA	-3,1248						
(S) ATGCGGTC	-3,6919		(S) GCCCGTGG	-2,7552						
(S) ATGCTGCG	-2,0576		(S) GCCCTCGG	-2,2837						
(S) ATGGCACG	-2,5455		(S) GCCGGTGC	-2,7933						
(S) ATGGCGGC	-2,3691		(S) GCCGTTTG	-3,0256						
(S) ATGGGGAC	-3,2032		(S) GCCTGCGG	-2,7378						
(S) ATGGTTCG	-2,2134		(S) GCCTGCGG	-3,0040						
(S) ATGTAGGA	-2,4373		(S) GCCTGGGG	-2,4517						
(S) ATGTGGGC	-2,6647		(S) GCCTGTGG	-3,3099						
(S) ATTCAGGC	-2,0618		(S) GCGCTTTC	-2,4893						
(S) ATTCCGGC	-2,1491		(S) GCGGGAAA	-2,5597						
(S) ATTCTGGG	-3,5348		(S) GCGGGGAA	-3,3319						
(S) ATTGGCGC	-2,2858		(S) GCGGGTGG	-4,0520						
(S) ATTGTGGC	-2,3164		(S) GCTATCGA	-3,0000						
(S) ATTTGCGC	-2,1022		(S) GCTCCGGG	-2,4535						
(S) ATTTTGCC	-2,1174		(S) GCTCGCGC	-2,5201						
(S) CAAAGGTG	-2,3437		(S) GCTCGCTA	-2,8526						
(S) CAACAGTG	-2,3770		(S) GCTCGTGC	-2,4295						
(S) CAACCGGC	-2,2955		(S) GCTCGTTA	-3,6455						
(S) CAACGCGC	-2,1736		(S) GCTCGTTG	-3,0626						
(S) CAACGGAG	-2,5207		(S) GCTGCTGC	-2,2845						
(S) CAACGTGG	-2,0796		(S) GCTGGGGC	-3,6772						
(S) CAAGCAGG	-2,3596		(S) GCTGTGGC	-2,4279						
(S) CAAGCGTC	-2,3507		(S) GCTTAGGC	-3,3486						
(S) CAAGGCTG	-2,7159		(S) GCTTGGGC	-3,2622						
(S) CAAGTTGG	-2,0871		(S) GCTTGTAC	-2,2739						
(S) CACCCGCG	-2,6818		(S) GGAACCTG	-2,3396						
(S) CACCGTAC	-2,1638		(S) GGACGTCC	-2,4871						
(S) CACCTTAG	-2,4012		(S) GGAGGCAG	-2,6399						
(S) CACGCATG	-2,4261		(S) GGAGGGGG	-2,4801						



(S) CACGCGGC	-2,8262		(S) GGAGTGGG	-4,0028					
(S) CACGGGTG	-2,6958		(S) GGATGCGC	-2,4538					
(S) CACGGTAG	-2,4458		(S) GGATGTCG	-2,2931					
(S) CACGGTTC	-3,7295		(S) GGATGTGC	-2,3703					
(S) CACGTGTC	-2,5046		(S) GGATGTGG	-3,7930					
(S) CACTCGGC	-2,2568		(S) GGCACCTG	-3,0604					
(S) CAGATTGC	-2,1909		(S) GGCAGCCA	-2,9646					
(S) CAGCCGCG	-2,5414		(S) GGCAGTCC	-2,2400					
(S) CAGCCGTC	-2,1150		(S) GGCAGTCG	-3,1813					
(S) CAGCGGCG	-3,3745		(S) GGCAGTGG	-2,8681					
(S) CAGGCATA	-2,0555		(S) GGCAGTTG	-2,4298					
(S) CAGGCGCC	-2,2953		(S) GGCATCGG	-2,7688					
(S) CAGGCTCG	-3,0415		(S) GGCATTGG	-2,4238					
(S) CAGGCTTG	-2,7617		(S) GGCCATGG	-2,5007					
(S) CAGGGACC	-2,3770		(S) GGCCCGGG	-3,9862					
(S) CAGGGCCG	-3,4094		(S) GGCCCTCG	-2,6107					
(S) CAGGGCGA	-3,3306		(S) GGCCCTTC	-3,6102					
(S) CAGGGGTA	-2,2936		(S) GGCCGGCG	-2,3241					
(S) CAGGGTGC	-2,7391		(S) GGCCGTCC	-2,5612					
(S) CAGGGTTA	-2,0892		(S) GGCCGTGG	-2,3747					
(S) CAGGGTTG	-2,5446		(S) GGCCTTAC	-2,4717					
(S) CAGGTCTA	-2,1988		(S) GGCCTTGC	-2,8134					
(S) CAGGTGGA	-2,2091		(S) GGC CGTA	-2,3322					
(S) CAGGTGGG	-2,2259		(S) GGC GCTG	-2,4423					
(S) CAGGTGTA	-3,2847		(S) GGC GGGAG	-2,8067					
(S) CAGGTAA	-2,3793		(S) GGC GGGCG	-2,2711					
(S) CAGGTTTC	-2,6729		(S) GGCTCGGG	-2,8353					
(S) CAGTATGG	-3,3823		(S) GGCTGCAC	-2,4041					
(S) CAGTCGAG	-2,1087		(S) GGCTGCGA	-2,3293					
(S) CAGTCTTC	-2,7182		(S) GGCTGTAG	-3,0964					
(S) CAGTCTTG	-3,0982		(S) GGCTTTCA	-2,2566					
(S) CATAGCGC	-2,5171		(S) GGG AAGGG	-2,7557					
(S) CATAGTGC	-2,5772		(S) GGG AATGG	-2,3846					
(S) CATATGTG	-2,5211		(S) GGGACGGC	-2,5189					
(S) CATCCCGG	-2,0782		(S) GGGACGTC	-2,9739					
(S) CATCGGTA	-2,6235		(S) GGGAGCAA	-3,3561					
(S) CATCGGTG	-2,2236		(S) GGGAGGGC	-2,6274					
(S) CATGCAGC	-2,0992		(S) GGGAGTGC	-2,3926					
(S) CATGCGGA	-2,2622		(S) GGGAGTGG	-2,5341					
(S) CATGGGAC	-2,0577		(S) GGGCAAGC	-2,3144					
(S) CATGGGTA	-3,3450		(S) GGGCATGC	-2,7085					
(S) CATGTGCG	-2,4389		(S) GGGCATTG	-2,3909					
(S) CATTTGTC	-2,6603		(S) GGGCCTGC	-2,6100					
(S) CATTTTCG	-2,8492		(S) GGGCGCGC	-2,3007					
(S) CCACAGGG	-2,1009		(S) GGGCGCGG	-3,4058					
(S) CCACGCGC	-2,1169		(S) GGGCGGGC	-2,8938					
(S) CCACGCGC	-2,3881		(S) GGGCGGTC	-3,3126					
(S) CCAGGATG	-2,1309		(S) GGGCTCGG	-2,6849					
(S) CCAGGTTA	-2,1846		(S) GGGGACGA	-3,8654					
(S) CCATGCGG	-2,1688		(S) GGGGACGC	-3,4911					
(S) CCATGGTC	-2,5050		(S) GGGGACGG	-3,0719					
(S) CCATGTTA	-2,1159		(S) GGGGAGGA	-2,9587					
(S) CCATTGGC	-2,5632		(S) GGGGATGC	-2,7796					
(S) CCATGTC	-2,5052		(S) GGGGAGG	-2,6555					
(S) CCCCAGAA	-2,2044		(S) GGGGCCAG	-3,1024					
(S) CCGCGCGA	-2,5207		(S) GGGGCGTA	-2,7610					
(S) CCGCGGGC	-2,6624		(S) GGGGCGTC	-3,8581					
(S) CCGGCGCC	-2,6809		(S) GGGGGACA	-4,2118					
(S) CCGGGGTG	-2,1139		(S) GGGGGATA	-2,7220					
(S) CCGGTGCG	-2,1352		(S) GGGGGCAA	-2,7535					
(S) CCGGTGTA	-2,1366		(S) GGGGGCCC	-2,4769					
(S) CCCTAGGG	-2,7072		(S) GGGGGCCG	-2,2683					
(S) CCCTGGGC	-2,8156		(S) GGGGGCTG	-2,8268					
(S) CCCTGGTC	-2,2786		(S) GGGGGGTA	-2,6802					
(S) CCCTGTTG	-3,0921		(S) GGGGGTAG	-2,5196					
(S) CCCTTCAG	-2,1827		(S) GGGGGTTA	-2,4944					
(S) CCGAGGGA	-2,0947		(S) GGGGTATG	-2,3945					
(S) CCGAGTTC	-2,8044		(S) GGGGTCCC	-2,4324					
(S) CCGATCGG	-2,2235		(S) GGGGTGCG	-2,3411					
(S) CCGCGCTC	-2,1905		(S) GGGGTGGC	-3,8469					
(S) CCGCGTTC	-2,6530		(S) GGGGTGGG	-2,5377					
(S) CCGGCGTG	-2,0686		(S) GGGGTGTC	-2,4282					
(S) CCGGGGTA	-3,0002		(S) GGGTCTCA	-3,0403					
(S) CCGGGGTC	-2,4138		(S) GGGTCTTA	-2,3206					
(S) CCGGTTTG	-2,9467		(S) GGGTGAGG	-2,8134					
(S) CCGTCTCG	-2,0973		(S) GGGTGGGG	-2,3944					
(S) CCGTCTTA	-2,4182		(S) GGGTTCGG	-2,4116					
(S) CCGTGCGC	-2,1419		(S) GGGTTGGC	-2,5413					
(S) CCGTGGCG	-3,3522		(S) GGGTTTGA	-2,5763					
(S) CCGTGTGA	-2,0897		(S) GGGTTTGC	-3,0488					
(S) CCGTGTTG	-2,2131		(S) GGTACAGG	-2,6414					

(S) CCGTTACA	-2,3394		(S) GGTACGGC	-4,6773						
(S) CCGTTGGA	-2,3648		(S) GGTACGGG	-2,3372						
(S) CCGTTGTG	-2,3677		(S) GGTACTCG	-2,3437						
(S) CCTAATTG	-2,1432		(S) GGTAGCAA	-2,6210						
(S) CCTAGGTG	-2,4604		(S) GGTAGCGA	-2,3039						
(S) CCTCAGTG	-2,0637		(S) GGTAGGGA	-4,4379						
(S) CCTCGTGG	-2,2842		(S) GGTAGTCG	-2,5837						
(S) CCTCTGGC	-2,4374		(S) GGTAGTTA	-2,3518						
(S) CCTGATCG	-2,3234		(S) GGTATTCT	-2,2470						
(S) CCTGGCCC	-2,1046		(S) GGTCAAAC	-2,5468						
(S) CCTGGGTA	-2,5496		(S) GGTCAAGC	-3,2546						
(S) CCTTCTTC	-2,6189		(S) GGTCCCTC	-2,8753						
(S) CCTTGTCT	-2,3203		(S) GGTCCCTG	-2,4470						
(S) CCTTGGAA	-2,4140		(S) GGTGAGCG	-3,1027						
(S) CCTTGGCC	-2,1736		(S) GGTGCGAG	-2,4113						
(S) CCTTGGTC	-3,9338		(S) GGTGCGGA	-2,9881						
(S) CCTTTGCG	-2,3621		(S) GGTGCGGG	-2,2956						
(S) CGAAGGGG	-3,0200		(S) GGTGCGTA	-2,9207						
(S) CGAAGGTG	-3,0295		(S) GGTCTGTG	-2,5003						
(S) CGAATGGG	-2,3122		(S) GGTCTTTC	-2,2667						
(S) CGAGGGCG	-2,2309		(S) GGTGAGCG	-2,3458						
(S) CGAGTTGG	-2,6070		(S) GGTGCGGC	-4,7879						
(S) CGCAAGGC	-2,4487		(S) GGTGCGGG	-2,7183						
(S) CGCACCTG	-2,4820		(S) GGTGCTGA	-2,8809						
(S) CGCACGGG	-2,6054		(S) GGTGCGAG	-2,7787						
(S) CGCAGGAG	-2,1735		(S) GGTGGGAC	-4,0040						
(S) CGCAGGTC	-2,7008		(S) GGTGGGGG	-2,8010						
(S) CGCAGTAG	-2,1177		(S) GGTGGGTA	-4,6048						
(S) CGCATGCA	-3,2039		(S) GGTGGGTC	-3,4674						
(S) CGCCACTA	-2,1370		(S) GGTGTCCG	-4,2817						
(S) CGCCCAAG	-2,5499		(S) GGTGTCTC	-2,8231						
(S) CGCCCAAG	-2,8862		(S) GGTGTCTG	-2,3638						
(S) CGCCCCCG	-2,1467		(S) GGTGTGCC	-2,4116						
(S) CGCCCCGC	-3,1831		(S) GGTGTGTC	-2,2447						
(S) CGCCCTGG	-2,7451		(S) GGTGTTGC	-2,6357						
(S) CGCCGCGC	-2,5411		(S) GGTGTTTA	-2,6319						
(S) CGCCGGCG	-2,0601		(S) GGTTCACC	-2,5377						
(S) CGCCGGTA	-2,1572		(S) GGTTCGCG	-2,4300						
(S) CGCCTGCA	-2,3853		(S) GGTTCGAA	-2,3617						
(S) CGCCTTCG	-2,3733		(S) GGTTCGTC	-4,2692						
(S) CGCCTTTG	-2,1658		(S) GGTTCGCG	-2,3517						
(S) CGCCAGTG	-2,8046		(S) GGTTCCTA	-2,6589						
(S) CGCGCCTC	-2,1402		(S) GGTTCGTC	-2,4456						
(S) CGCGCGGC	-3,2966		(S) GTACCAGC	-3,3859						
(S) CGCGCGTG	-2,7657		(S) GTACCGGG	-2,4261						
(S) CGCGGCAG	-2,1843		(S) GTACCGGG	-3,3640						
(S) CGCGGCGG	-2,6332		(S) GTACGTGC	-2,4885						
(S) CGCGGGCG	-2,9772		(S) GTACGTGA	-3,0098						
(S) CGCGGTAA	-2,3365		(S) GTATGGCC	-2,8710						
(S) CGCGTACG	-3,2609		(S) GTATGGGC	-2,7236						
(S) CGCTCAGG	-2,3909		(S) GTATTTTG	-2,2350						
(S) CGCTCGTG	-2,0607		(S) GTCAACGG	-2,7601						
(S) CGCTCTTC	-2,1167		(S) GTCAATGC	-2,7539						
(S) CGCTGTAA	-2,1863		(S) GTCAGCAG	-2,2345						
(S) CGCTTCAC	-2,1711		(S) GTCAGCGC	-2,9947						
(S) CGCTTCGG	-2,4292		(S) GTCAGCTC	-3,3281						
(S) CGCTTCGA	-3,1468		(S) GTCAGTGC	-3,3336						
(S) CGGAAGTG	-2,5124		(S) GTCATCTA	-2,4319						
(S) CGGAGCTC	-2,1988		(S) GTCCAAGC	-2,3060						
(S) CGGCCTCC	-2,4889		(S) GTCCAGGG	-2,4575						
(S) CGGCCTGC	-3,3900		(S) GTCCATGC	-2,7573						
(S) CGGGATGG	-2,5263		(S) GTCCCAAG	-2,2762						
(S) CGGGCACC	-2,4337		(S) GTCCCGGG	-2,6233						
(S) CGGGGCAA	-3,0694		(S) GTCCCTAC	-2,5436						
(S) CGGGGCAG	-2,2527		(S) GTCCGCGC	-2,4520						
(S) CGGGTCGA	-2,5106		(S) GTCCGGGG	-2,6605						
(S) CGGGTGCC	-2,5974		(S) GTCCTGGC	-2,3919						
(S) CGGGTGTG	-2,0801		(S) GTCGGCAC	-2,2868						
(S) CGGTAAAG	-2,2091		(S) GTCGTAC	-2,2354						
(S) CGGTACGC	-2,0745		(S) GTCGTGGG	-2,2785						
(S) CGGTGCGC	-2,2152		(S) GTCGTGTA	-3,0650						
(S) CGGTGCAC	-2,2379		(S) GTCGTTGC	-2,3846						
(S) CGGTGCGA	-2,6167		(S) GTCGTAAG	-2,8640						
(S) CGGTGGAA	-3,4468		(S) GTCGAGG	-2,4787						
(S) CGGTGGGC	-2,6131		(S) GTCGTCAC	-2,9781						
(S) CGTAAAGG	-2,3437		(S) GTCGCGC	-2,6928						
(S) CGTAAAGT	-2,5503		(S) GTCGGGGC	-2,9843						
(S) CGTAGGAA	-2,6206		(S) GTCGGGGG	-3,8523						
(S) CGTAGTGC	-2,1494		(S) GTCGTGTC	-2,3907						
(S) CGTATATG	-2,1420		(S) GTCCTACC	-2,3297						
(S) CGTCCCCA	-2,3779		(S) GTCCTCCG	-2,5477						

(S) CGTCCGAC	-2,3368		(S) GTGCACGG	-2,7890					
(S) CGTCGTTT	-2,1680		(S) GTGCCGCG	-2,4931					
(S) CGTCTGCG	-2,3635		(S) GTGCCTTA	-3,4501					
(S) CGTCTGGC	-3,7514		(S) GTGCGAGG	-2,8521					
(S) CGTGATGG	-2,3711		(S) GTGCGATC	-2,4688					
(S) CGTGCAAG	-2,0705		(S) GTGGAGAA	-2,7815					
(S) CGTGCCCTG	-2,1203		(S) GTGGAGGC	-2,7331					
(S) CGTGCGAG	-2,0576		(S) GTGGCAGC	-2,4393					
(S) CGTGCTGC	-2,4947		(S) GTGGCCTC	-2,3985					
(S) CGTGGCCC	-2,2808		(S) GTGGCGCG	-2,3029					
(S) CGTGGGTG	-2,5642		(S) GTGGCTTG	-2,2829					
(S) CGTGTGAC	-2,4667		(S) GTGGGACG	-2,6986					
(S) CGTGTGCA	-2,5498		(S) GTGGGCGG	-3,7888					
(S) CGTTCCGC	-2,9765		(S) GTGGGCGC	-2,4068					
(S) CGTTCCGG	-2,7270		(S) GTGGGGGC	-2,7278					
(S) CGTTCCGA	-2,9478		(S) GTGGGTGA	-2,2940					
(S) CGTTCTGTG	-2,2487		(S) GTGGGTGG	-2,5494					
(S) CGTTGCAG	-3,5063		(S) GTGGTCCG	-3,0458					
(S) CGTTGCGC	-2,6417		(S) GTGGTGGC	-2,5055					
(S) CGTTGCTC	-2,5257		(S) GTGTACGC	-2,3826					
(S) CGTTGGCG	-2,0591		(S) GTGTGATC	-2,2401					
(S) CGTTGTAC	-2,0999		(S) GTGTTCTA	-3,1756					
(S) CGTTGTAG	-2,6428		(S) GTGTTGTG	-2,4457					
(S) CGTTTCGC	-3,3791		(S) GTTAGCGA	-2,5063					
(S) CGTTTGGC	-2,6498		(S) GTTAGCGC	-2,5067					
(S) CGTTTTCG	-2,1508		(S) GTTAGCGG	-2,5644					
(S) CTACTGTG	-2,5780		(S) GTTAGGGG	-2,3531					
(S) CTATCGGA	-2,2656		(S) GTTAGTGC	-2,3566					
(S) CTATCGGC	-2,5982		(S) GTTATCGC	-2,9000					
(S) CTATGTCG	-2,6303		(S) GTTATCGG	-2,9653					
(S) CTCACGGG	-2,8533		(S) GTTCCATG	-2,6915					
(S) CTCACCTTA	-2,0992		(S) GTTCGCGC	-2,4206					
(S) CTCAGAGG	-2,1805		(S) GTTCGCTG	-2,4108					
(S) CTCAGATG	-2,8879		(S) GTTCGGGC	-3,2030					
(S) CTCAGGCG	-2,0566		(S) GTTCGGGG	-2,5592					
(S) CTCAGGGG	-3,2511		(S) GTTCTTCG	-2,7214					
(S) CTCAGTTG	-2,2355		(S) GTTGCTCG	-2,8859					
(S) CTCGCCGA	-2,3493		(S) GTTGCTGC	-2,6183					
(S) CTCGCGGC	-3,3507		(S) GTTGCTGG	-2,8686					
(S) CTCGGGTG	-2,3806		(S) GTTGGGAG	-2,5707					
(S) CTCGGTAG	-2,1189		(S) GTTGGGCC	-4,0989					
(S) CTCCTGGG	-2,2765		(S) GTTGGGCG	-4,0566					
(S) CTCGATGC	-2,5928		(S) GTTGGGGA	-2,9788					
(S) CTCGCGCC	-2,7486		(S) GTTGGGGG	-2,3925					
(S) CTCGGAAG	-2,1688		(S) GTTGGTCG	-2,3851					
(S) CTCGGGTA	-2,4316		(S) GTTTAGGC	-2,2699					
(S) CTCGGTGG	-2,3606		(S) GTTTCGGG	-2,5541					
(S) CTCGTGCA	-2,1403		(S) GTTTGGAA	-3,3488					
(S) CTCGTGGC	-2,1199		(S) GTTTTCAG	-2,3783					
(S) CTCGTGTG	-2,0745		(S) TAAGGGGG	-3,4065					
(S) CTCTCAGG	-2,6882		(S) TACGGGGC	-2,3547					
(S) CTCTCGGG	-2,2080		(S) TACTGGAG	-2,3725					
(S) CTCTGCCC	-2,5301		(S) TACTGGCG	-2,2746					
(S) CTCTGCGA	-2,2282		(S) TACTGGGG	-2,7569					
(S) CTCTGGTC	-2,9526		(S) TAGCGGGG	-2,8447					
(S) CTGCACGC	-2,1145		(S) TAGGGTGC	-2,5668					
(S) CTGCACGG	-2,1874		(S) TAGTGTGC	-3,8586					
(S) CTGCCCGC	-2,2878		(S) TATGGGCC	-2,8764					
(S) CTGCCGCC	-2,5341		(S) TATGGGGA	-2,9300					
(S) CTGCGTCG	-2,2623		(S) TATTTCGC	-2,6232					
(S) CTGCTCGA	-2,4257		(S) TCATCGGA	-2,4160					
(S) CTGGAATG	-2,3039		(S) TCATTGGG	-3,6832					
(S) CTGGCAAG	-2,0904		(S) TCCAGCCG	-2,7223					
(S) CTGGCCTG	-2,4882		(S) TCCATCGG	-2,9499					
(S) CTGGCGGA	-2,5583		(S) TCCCGGGA	-2,7421					
(S) CTGGCTCG	-2,2777		(S) TCCGTTAG	-2,2937					
(S) CTGGGAAG	-2,1319		(S) TCCGTTGC	-2,8272					
(S) CTGGGGTC	-2,2926		(S) TCCTGCGC	-2,4433					
(S) CTGGGGTA	-2,8602		(S) TCTACTTG	-2,3696					
(S) CTGGTCGG	-2,7550		(S) TCTAGCGC	-2,3060					
(S) CTGGTGCG	-2,6843		(S) TCTCCCGG	-2,2455					
(S) CTGGTTGA	-2,0627		(S) TCTGTAC	-2,5123					
(S) CTGTAAGC	-2,0837		(S) TCTGTGTA	-3,6764					
(S) CTGTAATG	-2,1306		(S) TCTGTGAC	-2,3229					
(S) CTGTACTG	-2,1765		(S) TCITTTACG	-2,6351					
(S) CTGTAGCA	-2,1306		(S) TGAAGGGA	-3,5420					
(S) CTGTCCCG	-2,3782		(S) TGACCTTC	-3,0090					
(S) CTGTCTAA	-2,4268		(S) TGACGCGC	-2,5061					
(S) CTGTCTTG	-2,3143		(S) TGACGGGC	-2,3215					
(S) CTGTGCCC	-2,0757		(S) TGACGGGG	-2,3937					
(S) CTGTGGAG	-3,5871		(S) TGAGGCGC	-2,6658					

(S) CTGTGGCG	-2,2493		(S) TGAGTCGG	-2,4344					
(S) CTTAATGC	-2,0663		(S) TGATGGGC	-4,4116					
(S) CTTATCGG	-2,6600		(S) TGCAATGC	-2,3315					
(S) CTTACAGA	-2,5720		(S) TGCACCTG	-3,6376					
(S) CTTCCGGA	-2,5040		(S) TGCAGCAG	-2,6220					
(S) CTTCCCGG	-2,2708		(S) TGCAGCCC	-2,2817					
(S) CTTCCGGC	-2,4457		(S) TGCAGTGC	-2,4480					
(S) CTTCCGGA	-2,6157		(S) TGCAGTGG	-2,7131					
(S) CTTCCGTA	-4,0842		(S) TGCCCCGA	-2,7681					
(S) CTTCTGTC	-2,8807		(S) TGCCCCCTC	-2,8373					
(S) CTTGAGGC	-2,3984		(S) TGCCCCGTG	-2,4408					
(S) CTTGATCG	-3,1487		(S) TGCCGCGA	-2,2519					
(S) CTTGCATG	-2,2637		(S) TGCCGGGC	-2,3425					
(S) CTTGCGAA	-3,0760		(S) TGCCGTCC	-2,6671					
(S) CTTGCTAG	-2,1121		(S) TGCCGTGA	-2,6103					
(S) CTTGCTTG	-2,5241		(S) TGCCCTGC	-2,5317					
(S) CTTGGGTA	-2,6635		(S) TGCCCTGGG	-2,6552					
(S) CTTGTCGA	-2,1122		(S) TGCGCTAG	-3,1063					
(S) CTTGTGGC	-2,1093		(S) TGCGGGCG	-2,7968					
(S) CTTGTGTG	-3,0570		(S) TGCGGGGA	-3,3396					
(S) CTTGTTC	-2,0775		(S) TGCGGGTC	-3,7160					
(S) CTTTATCG	-2,1730		(S) TGCGGGTG	-2,9381					
(S) CTTTCATA	-2,3541		(S) TGCGGGTCG	-2,3323					
(S) CTTTGAGG	-2,6221		(S) TGCGGGTTA	-2,3235					
(S) CTTTGCAA	-2,3791		(S) TGCGGTTG	-2,2340					
(S) CTTTGCCA	-2,9267		(S) TGCGTTTA	-2,2870					
(S) CTTTGCGC	-2,6959		(S) TGCTCCGG	-3,2207					
(S) CTTTGGAG	-2,3707		(S) TGCTCGGC	-2,2550					
(S) CTTTGGTA	-2,3803		(S) TGCTCGGG	-3,5084					
(S) CTTTGTTA	-2,8074		(S) TGCTGCAC	-2,4189					
(S) CTTTGGG	-2,3550		(S) TGCTGCGC	-2,4888					
(S) GAATGGGA	-2,5754		(S) TGCTGTGC	-2,7502					
(S) GAATGGTG	-2,0628		(S) TGCTGTTC	-2,6410					
(S) GACCGGGC	-2,6667		(S) TGCTGTTG	-2,7052					
(S) GACCGTGC	-2,3588		(S) TGCTTCGA	-2,6260					
(S) GACGCCGG	-2,2104		(S) TGCTTGCG	-3,9728					
(S) GACGGGAA	-2,5207		(S) TGCTTGGA	-2,4262					
(S) GACTGGAA	-2,1803		(S) TGCTTGGG	-2,6470					
(S) GAGACGCG	-2,2195		(S) TGGAATGA	-2,8225					
(S) GAGCCGGG	-2,6118		(S) TGGAGCGG	-2,3964					
(S) GAGGGCAA	-2,8489		(S) TGGGCTTA	-2,3377					
(S) GAGGGTGC	-2,5233		(S) TGGGCTTC	-2,2689					
(S) GATAGGGG	-2,3598		(S) TGGGGCCA	-2,4297					
(S) GATCTCGC	-2,4056		(S) TGGGGGCC	-3,3507					
(S) GATGGGCA	-2,9874		(S) TGGGTCTA	-2,2522					
(S) GATGGGTA	-2,2566		(S) TGGTACGC	-2,4545					
(S) GATTGGAG	-2,1807		(S) TGGTCATC	-3,1635					
(S) GATTGGTA	-2,1159		(S) TGGTGCGG	-2,5571					
(S) GATTTGCA	-2,1253		(S) TGGTGTGG	-2,6244					
(S) GCAAGGGC	-3,1313		(S) TGGTTCGG	-2,6543					
(S) GCACGGGC	-2,5630		(S) TGGTTTGC	-3,2998					
(S) GCACGTGC	-2,4932		(S) TGTACGGC	-2,6082					
(S) GCAGGTGC	-2,9120		(S) TGTATTGC	-2,2324					
(S) GCATAGTC	-2,0645		(S) TGTACAGG	-2,4074					
(S) GCATGACC	-2,1249		(S) TGTCCCGG	-2,4472					
(S) GCATGGTC	-2,2689		(S) TGTCCGGA	-3,2366					
(S) GCCACTCG	-2,1222		(S) TGTCCGCG	-3,4164					
(S) GCCAGGCG	-2,9389		(S) TGTCCGGG	-2,8321					
(S) GCCAGTGC	-2,6385		(S) TGTCTGAC	-2,5097					
(S) GCCAGTGG	-2,2965		(S) TGTCTCGC	-2,8439					
(S) GCCATGTG	-2,1884		(S) TGTCCGGC	-2,4588					
(S) GCCCAGGG	-2,0782		(S) TGTGGGCC	-2,6408					
(S) GCCCATGG	-2,3988		(S) TGTGGGTG	-2,5553					
(S) GCCCGGGC	-2,4175		(S) TGTGTGCC	-2,2724					
(S) GCCCGTCG	-2,1016		(S) TGTGTTAC	-2,2859					
(S) GCCGGGAA	-2,2764		(S) TGTTAACC	-2,4286					
(S) GCCGGGCA	-2,1182		(S) TGTGGGGG	-2,6093					
(S) GCCTGGCG	-2,3891		(S) TGTTTGTG	-3,1518					
(S) GCCTGTAC	-2,4006		(S) TGTTTTAG	-3,0820					
(S) GCCTTGGG	-2,3394		(S) TTACGTGG	-2,5662					
(S) GCGCATGG	-2,6505		(S) TTAAGTCC	-2,9411					
(S) GCGGGCGG	-2,5632		(S) TTATCGTC	-3,3428					
(S) GCGGGGTG	-2,8107		(S) TTATCTGC	-2,3168					
(S) GCTCCACG	-2,3264		(S) TTATCTTG	-2,6007					
(S) GCTCTCGA	-2,6789		(S) TTATGGAA	-2,8239					
(S) GCTCTGGA	-2,7250		(S) TTATGTAC	-2,9895					
(S) GCTGCGGG	-3,9708		(S) TTATTATC	-2,6414					
(S) GCTGCAG	-2,6897		(S) TTATTCGC	-2,7285					
(S) GCTGGCGC	-3,4946		(S) TTATTTCG	-2,3367					
(S) GCTGGGTA	-2,5717		(S) TTCCTGGG	-3,1524					
(S) GCTGTGCC	-2,6799		(S) TTCAGGGG	-2,3219					

(S) GCTGTGGG	-2,5337		(S) TTCATCTG	-2,8067					
(S) GCTTAGCG	-2,8130		(S) TTCATGTG	-2,8697					
(S) GGAAGGGC	-2,7045		(S) TTCCTGGG	-2,6465					
(S) GGACGCAG	-2,1488		(S) TTCGTGTC	-2,5698					
(S) GGACGGTC	-2,3107		(S) TTCGTTTA	-2,5466					
(S) GGACGGTG	-2,2405		(S) TTCTATCA	-2,4888					
(S) GGAGCAGA	-2,3775		(S) TTCTGTCC	-2,2443					
(S) GGAGCGTG	-2,5081		(S) TTCTGTGG	-2,3763					
(S) GGAGGGCG	-3,4556		(S) TTCTTGGC	-2,4186					
(S) GGAGGGAC	-3,0903		(S) TTCTTTAA	-2,7137					
(S) GGAGGGCG	-2,3203		(S) TTCTTTGG	-2,5213					
(S) GGAGGTAC	-2,1272		(S) TTGCGGGC	-2,9709					
(S) GGAGGTGC	-2,1488		(S) TTGCTTGC	-2,7074					
(S) GGATCTGC	-2,8735		(S) TTGCTTTC	-2,4148					
(S) GGCACAGG	-3,5576		(S) TTGGGTTC	-3,8894					
(S) GGCACGTC	-2,5331		(S) TTGTACGC	-3,0255					
(S) GGCAGCAA	-2,0565		(S) TTGTATGC	-3,1491					
(S) GGCAGCGC	-2,1598		(S) TTGTGTTG	-2,2893					
(S) GGCAGGGG	-2,4650		(S) TTTAAGGG	-2,4279					
(S) GGCAGGTA	-3,8983		(S) TTTAAGTC	-2,7392					
(S) GGCAGGTG	-3,1302		(S) TTTAGCGG	-2,9142					
(S) GGCAGTTC	-2,1673		(S) TTTATCAC	-2,6812					
(S) GGCAATGC	-2,2733		(S) TTTGCGGG	-3,0870					
(S) GGCCAGGA	-2,6813		(S) TTTGATGC	-2,4001					
(S) GGCCCAAG	-2,4742		(S) TTTGGGTC	-2,3060					
(S) GGCCCGCC	-2,4824		(S) TTTGTTGC	-2,3469					
(S) GGCCGCGC	-2,7006		(S) TTTTACGC	-2,3120					
(S) GGCCGCTC	-2,4658		(S) TTTTACTC	-2,7599					
(S) GGCCGGTA	-2,8752		(S) TTTTCAGG	-2,3394					
(S) GGCCGTGA	-2,8290		(S) TTTTCCTG	-2,3677					
(S) GGCCTGCG	-2,1172		(S) TTTTCGTG	-2,3752					
(S) GGCGAGTC	-2,3770		(S) TTTTGTTA	-2,3453					
(S) GGCGCCGA	-2,0815		(S) TTTTGTTG	-2,2566					
(S) GGCGCGCC	-2,2463		(S) AAGTTTCG	-2,4130					
(S) GGCGGGCA	-2,6991		(S) ACTTGAGC	-2,4649					
(S) GGCGGTCC	-2,6034		(S) AGAACGGC	-3,2663					
(S) GGCGGTTG	-2,2504		(S) AGGAGGAG	-3,3031					
(S) GGCGTTGG	-2,2484		(S) AGTCAGTA	-2,9417					
(S) GGCTGGAC	-3,2308		(S) ATGCTCGC	-2,2579					
(S) GGCTGGCC	-2,1275		(S) CACGGGTA	-2,5734					
(S) GGCTTTGC	-2,8147		(S) CACTGGTC	-2,3839					
(S) GGGAAGGC	-2,2943		(S) CATTCGTG	-2,2985					
(S) GGGACACG	-2,0848		(S) CGAACTTC	-2,6838					
(S) GGGACATA	-2,3980		(S) CTCCTCTC	-3,0000					
(S) GGGACCGG	-2,5666		(S) CTGTTTTC	-2,5115					
(S) GGGAGTAG	-3,1859		(S) GATGCTCC	-2,4184					
(S) GGGACCAA	-2,2661		(S) GCAGGTAG	-2,9511					
(S) GGGCACGG	-2,5524		(S) GCCACAAG	-2,4311					
(S) GGGCACTG	-3,1794		(S) GCCCCGTG	-3,0499					
(S) GGGCAGGC	-2,1141		(S) GCCCTGTA	-2,3172					
(S) GGGCAGTG	-2,2631		(S) GCGTCGGC	-2,6538					
(S) GGGCATGG	-2,6641		(S) GCTGGTAC	-2,3995					
(S) GGGCGAGA	-2,1667		(S) GGAGCGGC	-2,4744					
(S) GGGCGAGG	-2,6395		(S) GGGACTAG	-2,3937					
(S) GGGCGGAC	-2,1097		(S) GGGGATCG	-3,3671					
(S) GGGCTCAG	-2,0631		(S) GTTGATCG	-2,7447					
(S) GGGCTGTG	-3,1432		(S) TCGAGTTC	-2,3739					
(S) GGGGAATC	-3,4824		(S) TCTGTATG	-2,8461					
(S) GGGGAGGC	-2,4829		(S) TGCCCCAAG	-2,3004					
(S) GGGGAGTG	-2,1147		(S) TGCCCCCTA	-2,2986					
(S) GGGGATTA	-2,3859		(S) TGTAGGTA	-2,5692					
(S) GGGGATTG	-2,1931		(S) TGTGATTC	-2,9142					
(S) GGGGCAAA	-2,1013		(S) TGTGTACC	-2,6431					
(S) GGGGGAAG	-2,9058		(S) TTCAATTG	-2,4894					
(S) GGGGGGAG	-2,0936		(S) TTCAGGTC	-2,2742					
(S) GGGGGTAC	-2,1475								
(S) GGGGGTCG	-2,7686								
(S) GGGGGTGA	-2,4880								
(S) GGGTAAGC	-2,6620								
(S) GGGTACGC	-2,5400								
(S) GGGTAGGC	-3,2151								
(S) GGGTATCA	-2,4254								
(S) GGGTATCG	-2,8085								
(S) GGGTATGG	-2,8550								
(S) GGGTGCAC	-3,6665								
(S) GGGTGCGA	-2,5013								
(S) GGGTGGGC	-2,0909								
(S) GGGTGTAC	-2,4427								
(S) GGGTTGTG	-2,0600								
(S) GGGTTTTC	-2,1643								
(S) GGTAAGGG	-2,5164								

[illegible]

[illegible]

[illegible]



(S) TTCCCCGG	-2,5137											
(S) TTCCCCGA	-2,4137											
(S) TTCCGCAC	-2,1221											
(S) TTCCGTCG	-3,0928											
(S) TTCCGTGG	-2,1836											
(S) TTCCCTCG	-2,5267											
(S) TTCCCTGC	-2,7227											
(S) TTCCGCCG	-2,5988											
(S) TTCCGTGC	-2,4593											
(S) TTCCGTGG	-2,4640											
(S) TTCTAATC	-2,1739											
(S) TTCTAGGG	-3,1799											
(S) TTCTAGTC	-2,1177											
(S) TTCTGTCG	-3,1226											
(S) TTGGGCGC	-2,5978											
(S) TTGGTCGC	-2,6088											
(S) TTGGTTGG	-2,4345											
(S) TTGTATGA	-2,1903											
(S) TTGTCTTC	-2,0961											
(S) TTGTTGCC	-2,2552											
(S) TTGTTGTG	-2,2072											
(S) TTTAGCAG	-2,9368											
(S) TTTAGGGA	-2,8523											
(S) TTTCCGCC	-2,6665											
(S) TTTCCGTA	-2,6359											
(S) TTTCTGA	-2,1975											
(S) TTTCTTA	-2,2953											
(S) TTTCTTCC	-2,2807											
(S) TTTGCATC	-2,2210											
(S) TTTGGCGC	-2,5346											
(S) TTTGGGTA	-2,5764											
(S) TTTGGTCG	-2,1541											
(S) TTTGGTGG	-2,0599											
(S) TTTGTTGA	-2,5720											
(S) TTTTACGG	-3,3687											
(S) TTTTCAGA	-2,3387											
(S) TTTTCCCC	-2,2688											
(S) TTTTCGCC	-3,0002											
(S) TTTTCTAG	-2,3032											
(S) TTTTCTTA	-2,4100											
(S) TTTTGAGG	-2,3024											
(S) TTTTGCCA	-2,1119											
(S) TTTTGGTG	-2,9199											
(S) TTTTCCCG	-2,3312											
(S) TTTTTCAT	-2,3313											

## XV. Validierung identifizierter SRMs

rHBBg-Plasmid	WA	HeLa				HuH7				SH-SY5Y			
		AI <sub>sk</sub>	E <sub>ein</sub>	SD	R <sup>2</sup>	AI <sub>sk</sub>	E <sub>ein</sub>	SD	R <sup>2</sup>	AI <sub>sk</sub>	E <sub>ein</sub>	SD	R <sup>2</sup>
WT (GGATCCTG)	0	1,15	96,64	2,96	0,957	3,94	98,02	1,57	0,893	6,61	97,73	1,66	0,901
Mu1 (GGGTCCTG)	1	0,24	86,58	3,59		-1,92	81,65	6,38		-2,09	73,95	2,84	
Mu2 (CCTTGCTG)	4	-2,32	74,65	5,86		-2,45	72,68	3,22		-0,42	83,81	4,12	
Mu3 (TAGACCTA)	5	2,92	99,34	4,82		2,65	98,27	2,13		2,28	93,34	4,89	
Mu4 (ACTCCGGG)	6	-0,38	85,23	2,89		-1,69	83,59	2,11		-1,28	79,95	5,89	
Mu5 (TTTAGGGA)	8	-2,85	69,14	4,99		-0,05	87,78	3,98		-2,45	73,48	6,15	

## XVI. Sequenz-Clusteranalyse

### • Enhancer-Sequenzen

Clusteranalyse		Enhancer-Datensatz (E)				Kontrolldatensatz (K)				E-K-Vergleich	
		Minimum	Maximum	Cluster-anzahl	IQR	Minimum	Maximum	Cluster-anzahl	IQR	Median-differenz	p-Wert (U-test)
HeLa	DL	1	43	81	10	1	36	82	8	2	0,7146
	PL	1	15	81	4	1	14	82	4	0	0,9572
HuH7	DL	1	40	47	10	1	36	48	10,25	0	0,9404
	PL	1	6	46	2	1	9	46	2	0	0,9353
SH-SY5Y	DL	1	9	35	3	1	13	35	2,5	0	0,9238
	PL	1	5	41	1	1	8	42	1	0	0,7073

### • Silencer-Sequenzen

Clusteranalyse		Enhancer-Datensatz (E)				Kontrolldatensatz (K)				E-K-Vergleich	
		Minimum	Maximum	Cluster-anzahl	IQR	Minimum	Maximum	Cluster-anzahl	IQR	Median-differenz	p-Wert (U-test)
HeLa	DL	1	33	75	9	1	31	79	11,5	1	0,7227
	PL	1	16	85	4	1	15	82	4	1	0,9987
HuH7	DL	1	35	71	13,5	1	31	76	13	1	0,9134
	PL	1	15	77	4	1	13	78	4	0	0,9666
SH-SY5Y	DL	1	45	31	14,5	1	36	32	17	0	0,9341
	PL	1	11	56	2	1	10	57	2	0	0,9315

## XVII. Annotation der identifizierten Enhancer- und Silencer-Sequenzen

### • KS-Test auf Normalverteilung der Scores der spleißregulatorischen Faktoren

spleiß-regulatorische Faktoren	p-Wert (t-Test)											
	HeLa				HuH7				SH-SY5Y			
	DL		PL		DL		PL		DL		PL	
	E	S	E	S	E	S	E	S	E	S	E	S
GC-Gehalt	0,6158	0,1108	0,1122	0,9274	0,7444	0,7146	0,9611	0,3552	0,6184	0,1414	0,4770	0,1143
Purin-Gehalt	0,6690	0,1277	0,3026	0,2139	0,4017	0,5992	0,1362	0,5850	0,0959	0,9332	0,6089	0,2717
WA	0,8225	0,7488	0,6094	0,3951	0,3200	0,3213	0,9496	0,5703	0,0365	0,6331	0,8285	0,5651
EF	0,7702	0,5820	0,7415	0,0736	0,4155	0,7483	0,5332	0,0286	0,1733	0,8260	0,5738	0,0223
hnRNP A1	0,8479	0,6132	0,5984	0,7626	0,3427	0,0551	0,2033	0,2373	0,2580	0,6298	0,1989	0,0261
hnRNP A/B	0,6977	0,7187	0,9968	0,3120	0,1964	0,9228	0,0341	0,8304	0,5630	0,8832	0,2184	0,5922
hnRNP H/F	0,5517	0,3798	0,3482	0,3245	0,9890	0,2299	0,5875	0,5481	0,4451	0,6759	0,0993	0,9028
hnRNP G	0,6376	0,4975	0,5964	0,8867	0,1925	0,9988	0,7458	0,4407	0,6921	0,6823	0,8876	0,7894
Tra2β	0,3821	0,3556	0,3507	0,0248	0,5036	0,9789	0,0977	0,6270	0,5240	0,4048	0,9012	0,4387
Tra2α	0,0654	0,3279	0,7179	0,7002	0,3072	0,9418	0,1608	0,3763	0,0731	0,5397	0,8148	0,9748
SC35	0,1411	0,0549	0,6869	0,7252	0,4905	0,4386	0,5819	0,8963	0,9248	0,4856	0,1628	0,6062
SF2/ASF	0,5357	0,8123	0,1447	0,0471	0,4689	0,0776	0,3706	0,2136	0,2483	0,5315	0,8172	0,8613
SRp20	0,0837	0,3153	0,2314	0,3206	0,9257	0,5130	0,5577	0,7575	0,5944	0,9561	0,0746	0,9312
SRp40	0,6233	0,8814	0,8298	0,2725	0,9607	0,1545	0,0363	0,8759	0,2724	0,2179	0,7684	0,5084

SRp55	0,4683	0,2596	0,9407	0,7557	0,4259	0,8421	0,1371	0,7783	0,7086	0,3607	0,1516	0,8575
9G8	0,6207	0,5531	0,6515	0,7524	0,3085	0,0920	0,9854	0,0957	0,6534	0,6197	0,7989	0,2532
SRp30c	0,8475	0,0433	0,1926	0,7812	0,3597	0,0442	0,7382	0,7292	0,6830	0,1472	0,9776	0,1884
CUG-BP	0,0204	0,5302	0,5429	0,1132	0,5921	0,7537	0,9279	0,4722	0,9538	0,1312	0,9402	0,4033
FOX1	0,5173	0,5060	0,9268	0,1255	0,1120	0,7610	0,8941	0,1632	0,6597	0,3188	0,5007	0,3868
MBNL1	0,1492	0,0609	0,4942	0,4199	0,2644	0,3039	0,8963	0,4915	0,8734	0,3392	0,6165	0,2329
NOVA1	0,3382	0,1735	0,8817	0,0397	0,9954	0,5091	0,4790	0,7808	0,3689	0,9865	0,3341	0,2953
YB1	0,9315	0,8555	0,1209	0,7902	0,2676	0,7888	0,2517	0,8495	0,7388	0,7569	0,4600	0,2050
PTB	0,5060	0,8950	0,4789	0,6928	0,9541	0,9961	0,1205	0,2204	0,9283	0,2990	0,5029	0,7347

• **t-Test der Scores der spleißregulatorischen Faktoren von Enhancern und Silencern**

spleiß-regulatorische Faktoren	HeLa						HuH7						SH-SY5Y					
	DL			PL			DL			PL			DL			PL		
	p-Wert (t-Test)	d	r	p-Wert (t-Test)	d	r	p-Wert (t-Test)	d	r	p-Wert (t-Test)	d	r	p-Wert (t-Test)	d	r	p-Wert (t-Test)	d	r
GC-Gehalt	4,77E-119	-1,25	0,53	1,07E-29	-0,97	0,44	9,70E-102	-1,38	0,54	1,03E-09	-1,08	0,43	3,28E-46	-3,30	0,81	7,00E-15	-1,19	0,48
Purin-Gehalt	1,18E-05	0,22	0,11	6,98E-04	-0,26	0,13	2,63E-06	0,34	0,16	0,01	-0,29	0,13	1,06E-06	-0,75	0,31	3,04E-03	-0,43	0,19
WA	1,40E-12	0,35	0,17	0,35	0,07	0,04	3,41E-08	0,33	0,15	0,05	0,21	0,09	2,02E-03	0,32	0,14	1,78E-03	0,45	0,20
EF	6,51E-72	0,97	0,44	2,31E-17	0,65	0,31	4,24E-47	1,32	0,52	2,52E-17	1,25	0,49	1,23E-11	1,32	0,49	4,22E-12	1,04	0,43
hnRNP A1	0,30	0,05	0,03	0,12	-0,12	0,06	5,11E-01	-0,04	0,02	0,02	-0,35	0,15	3,21E-08	-0,90	0,36	0,90	-0,02	0,01
hnRNP A/B	3,82E-39	0,67	0,32	2,64E-26	0,83	0,38	3,72E-26	0,89	0,38	7,77E-08	0,89	0,37	1,72E-13	1,50	0,54	5,09E-12	1,54	0,57
hnRNP H/F	1,21E-202	-1,75	0,66	2,03E-33	-0,97	0,44	3,74E-148	-1,74	0,63	4,66E-15	-0,89	0,37	2,23E-54	-3,78	0,85	3,71E-17	-1,30	0,51
hnRNP G	1,60E-23	-0,52	0,25	3,88E-04	-0,27	0,13	1,07E-19	-0,59	0,27	0,02	-0,31	0,14	2,41E-05	-0,44	0,19	0,33	-0,17	0,08
Tra2β	1,78E-08	-0,28	0,14	0,03	-0,18	0,09	0,06	-0,11	0,05	0,10	-0,30	0,13	1,22E-03	-0,34	0,14	0,05	-0,45	0,20
Tra2α	4,19E-07	0,25	0,13	4,70E-03	-0,21	0,11	3,66E-07	0,41	0,19	0,05	-0,22	0,10	0,51	0,07	0,03	0,01	-0,35	0,16
SC35	6,94E-10	-0,31	0,15	0,12	0,12	0,06	6,91E-10	-0,53	0,24	0,05	0,30	0,13	0,44	-0,14	0,06	0,77	0,04	0,02
SF2/ASF	3,36E-38	-0,66	0,31	1,56E-03	-0,25	0,12	5,69E-32	-0,90	0,38	0,76	0,05	0,02	7,30E-16	-0,87	0,35	0,02	-0,33	0,15
SRp20	0,19	0,06	0,03	8,16E-17	0,64	0,31	0,49	-0,05	0,02	1,59E-14	1,18	0,47	0,08	0,28	0,12	1,42E-09	0,90	0,38
SRp40	1,98E-27	-0,55	0,26	0,77	0,02	0,01	8,25E-31	-0,87	0,37	0,84	-0,02	0,01	5,33E-09	-0,91	0,36	0,35	-0,13	0,06
SRp55	0,29	0,05	0,03	6,18E-04	0,26	0,13	0,35	-0,06	0,03	2,39E-06	0,52	0,23	0,27	0,18	0,08	0,01	0,38	0,17
9G8	0,09	-0,08	0,04	1,18E-17	0,68	0,32	0,01	-0,16	0,08	1,84E-08	1,03	0,42	5,09E-05	-0,43	0,18	4,77E-07	0,74	0,32
SRp30c	2,09E-08	-0,28	0,14	2,95E-06	0,36	0,18	2,73E-06	-0,39	0,18	1,75E-05	0,74	0,32	0,00	-0,55	0,23	0,06	0,39	0,18
CUG-BP	3,55E-80	-0,99	0,44	9,25E-19	-0,69	0,32	9,88E-63	-1,32	0,52	1,07E-11	-1,19	0,47	1,30E-10	-1,13	0,43	1,85E-10	-1,29	0,51
FOX1	1,87E-33	-0,61	0,29	4,87E-10	-0,48	0,23	3,44E-25	-0,80	0,35	0,05	-0,21	0,09	1,20E-05	-0,46	0,19	0,01	-0,49	0,22
MBNL1	1,26E-54	-0,81	0,37	5,68E-24	-0,79	0,37	6,17E-43	-1,04	0,43	1,18E-11	-0,76	0,32	1,85E-03	-0,49	0,20	1,91E-08	-0,83	0,35
NOVA1	8,09E-14	0,37	0,18	9,44E-20	0,71	0,33	9,71E-10	0,47	0,21	1,69E-09	0,90	0,37	1,64E-11	0,72	0,29	5,75E-10	1,20	0,48
YB1	1,19E-118	1,26	0,53	6,75E-62	1,38	0,57	1,83E-78	1,77	0,63	2,32E-22	1,82	0,63	1,40E-36	2,88	0,77	2,37E-23	1,59	0,59
PTB	3,48E-23	-0,51	0,25	2,48E-03	-0,23	0,11	1,04E-26	-0,73	0,32	4,26E-04	-0,46	0,20	0,08	-0,24	0,10	0,18	-0,19	0,09

## XVIII. Multiple lineare Regressionsmodelle

### • DL- und PL-Analysen in HeLa-Zellen

Modell	R <sup>2</sup>	korrigiertes R <sup>2</sup>	Durbin-Watson-Statistik
DL HeLa	0,209	0,208	1,704
Koeffizienten	β	β standardisiert	p-Wert
Konstante	3,244	-	7,17E-195
GC-Gehalt	-1,232	-0,207	7,46E-206
EF	-0,009	-0,001	0,852
hnRNP A/B	-0,018	-0,002	0,75
hnRNP H/F	-1,861	-0,232	7,44E-294
hnRNP G	-0,61	-0,059	7,96E-28
SF2/ASF	-0,242	-0,022	2,18E-05
SRp40	-0,562	-0,053	5,94E-24
CUG-BP	-0,007	-0,001	0,891
FOX1	-0,619	-0,066	1,29E-38
MBNL1	-0,357	-0,04	2,88E-10
YB1	1,234	0,115	5,49E-94
PTB	-1,024	-0,16	4,45E-155

Modell	R <sup>2</sup>	korrigiertes R <sup>2</sup>	Durbin-Watson-Statistik
PL HeLa	0,096	0,095	1,593
Koeffizienten	β	β standardisiert	p-Wert
Konstante	-0,308	-	0,01
GC-Gehalt	-0,562	-0,131	1,42E-45
EF	0,478	0,081	5,22E-18
hnRNP A/B	0,386	0,05	7,83E-09
hnRNP H/F	-0,543	-0,065	2,37E-13
SRp20	0,312	0,055	1,38E-08
9G8	0,006	0,001	0,955
CUG-BP	-0,048	-0,008	0,497
MBNL1	-0,254	-0,039	7,91E-04
NOVA1	0,065	0,013	0,222
YB1	1,039	0,133	1,86E-32

### • DL- und PL-Analysen in HuH7-Zellen

Modell	R <sup>2</sup>	korrigiertes R <sup>2</sup>	Durbin-Watson-Statistik
DL HuH7	0,224	0,224	1,652
Koeffizienten	β	β standardisiert	p-Wert
Konstante	3,101	-	6,828E-105
GC-Gehalt	-1,478	-0,230	4,052E-195
EF	-0,035	-0,004	0,558
hnRNP A/B	0,028	0,002	0,692
hnRNP H/F	-1,882	-0,216	6,187E-204
hnRNP G	-0,887	-0,079	2,463E-39
SC35	0,419	0,032	3,827E-07
SF2/ASF	-0,208	-0,018	2,571E-03
SRp40	-0,569	-0,049	5,383E-17
CUG-BP	0,025	0,003	0,676
FOX1	-0,630	-0,061	3,921E-27
MBNL1	-0,494	-0,051	9,650E-13
YB1	1,685	0,145	4,584E-116
PTB	-1,223	-0,178	4,695E-149

Modell	R <sup>2</sup>	korrigiertes R <sup>2</sup>	Durbin-Watson-Statistik
PL HuH7	0,136	0,135	1,547
Koeffizienten	β	β standardisiert	p-Wert
Konstante	-1,379	-	5,058E-12
GC-Gehalt	-0,796	-0,142	9,494E-31
EF	0,844	0,108	3,389E-21
hnRNP H/F	-0,838	-0,078	1,064E-12
SRp20	0,554	0,075	5,902E-09
SRp55	0,680	0,074	3,775E-10
9G8	0,267	0,018	0,126
SRp30c	0,563	0,049	2,529E-05
CUG-BP	-0,099	-0,012	0,389
MBNL1	-0,507	-0,060	2,822E-05
NOVA1	-0,105	-0,016	0,216
YB1	1,427	0,137	1,424E-23

• DL- und PL-Analysen in SH-SY5Y-Zellen

Modell	R <sup>2</sup>	korrigiertes R <sup>2</sup>	Durbin-Watson-Statistik
DL SH-SY5Y	0,324	0,323	1,495
Koeffizienten	$\beta$	$\beta$ standardisiert	p-Wert
Konstante	2,381	-	3,484E-24
GC-Gehalt	-1,811	-0,261	1,960E-121
Purin-Gehalt	0,888	0,132	8,606E-27
EF	-0,364	-0,035	3,920E-04
hnRNP A1	-0,727	-0,065	3,830E-11
hnRNP A/B	0,490	0,036	1,038E-04
hnRNP H/F	-2,744	-0,291	1,281E-129
SF2/ASF	-0,193	-0,015	0,113
SRp40	-0,919	-0,073	8,526E-16
SRp30c	0,090	0,006	0,511
CUG-BP	-0,522	-0,058	1,329E-09
NOVA1	-0,406	-0,053	8,117E-07
YB1	2,918	0,227	3,861E-100

Modell	R <sup>2</sup>	korrigiertes R <sup>2</sup>	Durbin-Watson-Statistik
PL SH-SY5Y	0,219	0,217	1,523
Koeffizienten	$\beta$	$\beta$ standardisiert	p-Wert
Konstante	-1,041	-	9,050E-05
GC-Gehalt	-1,249	-0,208	1,957E-46
EF	0,761	0,092	3,333E-10
hnRNP A/B	0,799	0,074	6,164E-08
hnRNP H/F	-1,410	-0,124	3,991E-19
SRp20	1,053	0,134	1,621E-18
9G8	0,981	0,062	2,376E-05
CUG-BP	-0,274	-0,031	0,078
MBNL1	-0,165	-0,018	0,317
NOVA1	-0,070	-0,010	0,547
YB1	1,782	0,158	1,651E-19

---

## Erklärung

Hiermit erkläre ich, dass die vorliegende Dissertation, abgesehen von der Beratung durch meinen akademischen Lehrer, nach Inhalt und Form meine eigene Arbeit ist. Die Arbeit wurde weder vollständig noch in Teilen einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt. Die Abhandlung ist unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden. Ferner erkläre ich, dass ich noch keine früheren Promotionsversuche unternommen habe.

Kiel, den 10.10.2013

Frank Schumann

---

## Lebenslauf

### Persönliche Daten

Name: Frank Schumann  
Geburtstag: 02.01.1984  
Geburtsort: Gera  
Staatsangehörigkeit: deutsch  
Familienstand: ledig  
Kontaktadresse: fschumann@1med.uni-kiel.de

### Promotion

Seit 02/2009 Abteilung für Genomische Gastroenterologie der I. Medizinischen Klinik des Universitätsklinikums Schleswig-Holstein, Campus Kiel, unter Leitung von Prof. Dr. Jochen Hampe

### Studium

09/2004 - 12/2008 Studium der Pharma- und Biotechnologie an der Fachhochschule Jena mit dem Abschluss Dipl.-Ing. (FH)  
Studienschwerpunkte: Gentechnik, Protein-Engineering, Labor- und Analysentechnik

Praktische Diplomarbeit Abteilung für Genomanalyse des Leibniz-Institutes für Altersforschung – Fritz-Lipmann-Institut e.V. Jena, unter Leitung von PD Dr. Matthias Platzer  
Diplomthema: „Einsatz von Minigenen zur Analyse von Tandem-3'-Spleißmotiven in Säugerzellen“

10/2003 - 04/2004 Diplomstudium Chemie an der Friedrich-Schiller-Universität Jena

### Wehrdienst

07/2002 - 03/2003 Sanitätsdienst der Luftwaffe in Fürstenfeldbruck

### Schule

06/2002 Erwerb der Allgemeinen Hochschulreife am Albert-Schweitzer-Gymnasium in Gera

---

## Veröffentlichungen

- **Publikationen**

Szafranski K., Fritsch C., **Schumann F.**, Siebel L., Sinha R., Hampe J., Hiller M., Englert C., Huse K., Platzer M., 2013. Thousands of mammalian genes are co-regulated via a common alternative splicing mechanism. *Molecular and Cellular Biology*, Manuskript im August eingereicht.

Fritsch C., Herrmann A., Nothnagel M., Szafranski K., Huse K., **Schumann F.**, Schreiber S., Platzer M., Krawczak M., Hampe J., Brosch M., 2012. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Research*, 22: 2208–2218.

- **Posterpräsentationen**

**Schumann F.**, Brosch M., Herrmann A., Szafranski K., Huse K., Platzer M., Schreiber S., Hampe J., 2011. Deciphering the NAFLD associated splicing code. *International Conference on System Biology of Human Disease*, Boston, USA.



---

## Danksagung

Die vorliegende Arbeit entstand in der Abteilung für Genomische Gastroenterologie der I. Medizinischen Klinik des Universitätsklinikums Schleswig-Holstein, Campus Kiel mit freundlicher Genehmigung von Prof. Dr. Jochen Hampe. Ich bedanke mich bei ihm für die Bereitstellung dieses spannenden Themas, seinem Interesse am Fortgang der Arbeit und für die konstruktiven Diskussionen.

Ich danke Prof. Dr. Karin Krupinska für die Betreuung und Begutachtung der Arbeit seitens der Christian-Albrechts-Universität zu Kiel.

Mein herzlicher Dank gilt Dr. Mario Brosch für seine engagierte Unterstützung, besonders im Bereich experimenteller Arbeiten sowie seiner Hilfe bei der Überarbeitung meiner Manuskripte.

Großer Dank gilt allen Mitarbeiterinnen und Mitarbeitern der Arbeitsgruppe Hampe, ohne die diese Arbeit nicht möglich gewesen wäre. Ganz besonders möchte ich mich bei Alexander Herrmann für seine mathematische und bioinformatische Mitarbeit bedanken.

Ich danke Dr. Klaus Huse und Dr. Karol Szafranski für hilfreiche Telefonate rund um das Thema Spleißen. Weiterhin bin ich Oliver Vollrath für seine statistischen Lösungsansätze dankbar. Dr. Katja Klausz danke ich für das intensive Korrekturlesen dieser Arbeit.

Mein ganz besonderer Dank gilt abschließend meinen Eltern und meiner Freundin, die durch ihre einzigartige Unterstützung zum Gelingen dieser Arbeit beigetragen haben und denen diese Arbeit gewidmet ist, vielen Dank!